Détection des ruptures dans les processus causaux: Application aux débits du bassin versant de la Sanaga au Cameroun

William Charky Kengne

Ecole Doctorale de Sciences Mathématiques de Paris Centre

Ecole Nationale supérieure Polytechnique

THÈSE DE DOCTORAT

pour obtenir le garde de

DOCTEUR ES-SCIENCES
Spécialité : MATHÉMATIQUES APPLIQUÉES

présentée par

William Charky KENGNE

sous le titre

Détection des ruptures dans les processus causaux : Application aux débits du bassin versant de la Sanaga au Cameroun

Directeurs de Thèse

M. Jean-Marc BARDET, Professeur Université Paris 1
M. Henri GWET, Maitre de Conférences Université de Yaoundé I
M. Ouagnina HILI, Professeur INP Yamoussoukro

Soutenue publiquement le 03 mai 2012 à l'Université Paris 1 - Panthéon Sorbonne
devant le jury composé de :

Marc Lavielle, Professeur	INRIA	Président
Paul Doukhan, Professeur	Université Cergy Pontoise	Rapporteur
Eric Moulines, Professeur	Télécom ParisTech	Rapporteur
Tadjuidje K. Joseph, Associate	University Kaiserslautern	Examinateur
Jean-Marc Bardet, Professeur	Université Paris 1	Directeur de Thèse
Konstantinos F., Professor	University of Cyprus	Invité
Dag Tjøstheim, Professor	University of Bergen	Invité
José R. Leon , Profesor	Universidad C. Venezuela	Invité

A ma mère

Remerciements

Mes remerciements vont tout d'abord au Professeur Jean-Marc Bardet, mon directeur de thèse qui est l'initiateur de ce travail. Je lui exprime toute ma gratitude pour son engagement, son soutien, ses conseils et sa disponibilité... Son dynamisme, sa rigueur et ses qualités scientifiques ont été essentiels pour l'accomplissement de ce travail. Il m'a aussi beaucoup aidé dans la recherche de financements et dans la gestion administrative de la cotutelle entre Paris et Yaoundé.

Je remercie également le Professeur Henri Gwet, le co-encadreur de ce travail qui a accepté de me prendre en thèse. Il est le responsable du master de Yaoundé dont je suis un pur produit. C'est avec lui que j'ai fait mes premiers pas dans la recherche depuis mon mémoire de master. Il a été toujours disponible sur le plan scientifique et administratif.

Je remercie le Professeur Ouagnina Hili le co-directeur de cette thèse pour avoir accepté de travailler avec moi. J'ai bénéficié de son assistance à distance. Suite à la crise postélectorale de 2010 en Côte d'Ivoire, on n'a pas pu effectuer le travail qu'il avait prévu.

A tous mes trois directeurs de thèse, je leur suis profondément reconnaissant de m'avoir donné l'opportunité de travailler avec eux et d'effectuer ce travail.

Je remercie grandement Olivier Wintenberger, avec qui on a beaucoup collaboré dans la première partie de cette thèse. Il m'a aussi aidé et m'a assisté dans la suite.

Mes remerciements vont aussi au Professeur Didier Dacunha-Castelle et au Professeur Elisabeth Gassiat qui ont su (avec Jean-Marc Bardet) piloter le projet STAFAV avec succès. Ils ont été déterminants pour le financement de cette thèse. Ils m'ont apporté toute leur assistance sur le plan scientifique, administratif et logistique.

Je tiens à remercier le Professeur Marie Cottrell, la Directrice du laboratoire SAMM au sein duquel j'ai travaillé à Paris. Elle a été toujours à mes côtés pour se rassurer que tout allait bien et que j'avais tout ce qu'il fallait pour travailler. Elle m'a assisté dans mes procédures administratives et ne s'est pas fatiguée de signer des attestations pour moi.

Je remercie chaleureusement tous les membres du laboratoire SAMM pour leur accueil et leur esprit d'équipe. Ils m'ont accompagné dans mes travaux et ont su créer un environnement sain pour rendre le travail agréable au SAMM. Je remercie tout particulièrement les doctorants du SAMM comme Omar, Fania, Anthikos, Béchir, Mouhamad, Ibrahima,... pour leur solidarité. Je remercie aussi toute l'équipe du laboratoire LIMSS de Yaoundé en particulier, Monsieur Ndong Nguema, Cyprien, Patrice, Jules pour toute leur collaboration.

Je remercie l'Agence Universitaire de la Francophonie (AUF) et le projet EDULINK ACP-EU qui ont fortement contribué au financement de cette thèse.

Un grand merci au Professeur Paul Doukhan avec qui j'ai eu plusieurs discussions qui ont contribué à améliorer ce travail. Il m'a aussi invité à plusieurs conférences qu'il a organisées. Cela m'a permis de rencontrer et de discuter avec des grands experts de la détection de rupture.

Je tiens à exprimer ma reconnaissance à Madame Marie Granier pour toute l'assistance logistique qu'elle m'a apportée. Elle m'a fait entièrement confiance quand je suis arrivé à Paris et a accepté de m'héberger chez elle.

Je remercie aussi tous les amis et amies pour tous leurs encouragements. Ils ont contribué chacun à sa manière à la réalisation de ce travail.

Enfin, je voudrais remercier tous les membres de ma famille et de la famille Tchuenté pour leur soutien et leur accompagnement permanent. Ils ont été à mes côtés pendant ces années et m'ont apporté tout leur amour.

Résumé

Résumé

Cette thèse porte sur la détection de rupture dans les processus causaux avec application aux débits du bassin versant de la Sanaga. Nous considérons une classe semi-paramétrique de modèles causaux contenant des processus classique tel que l'AR, ARCH, TARCH.

Le chapitre 1 est une synthèse des travaux. Il présente le modèle avec des exemples et donne les principaux résultats obtenus aux chapitres 2, 3,4.

Le chapitre 2 porte sur la détection off-line de ruptures multiples en utilisant un critère de vraisemblance pénalisée. Le nombre de rupture, les instants de rupture et les paramètres du modèle sur chaque segment sont inconnus. Ils sont estimés par maximisation d'un contraste construit à partir des quasi-vraisemblances et pénalisées par le nombre de ruptures. Nous donnons les choix possibles du paramètre de pénalité et montrons que les estimateurs des paramètres du modèle sont consistants avec des vitesses optimales. Pour des applications pratiques, un estimateur adaptatif du paramètre de pénalité basé sur l'heuristique de la pente est proposé. La programmation dynamique est utilisée pour réduire le coût numérique des opérations, celui-ci est désormais de l'ordre de $\mathcal{O}(n^2)$. Des comparaisons faites avec des résultats existants montrent que notre procédure est plus stable et plus robuste.

Le chapitre 3 porte toujours sur la détection off-line de ruptures multiples, mais cette fois en utilisant une procédure de test. Nous avons construit une nouvelle procédure qui, combinée avec un algorithme de type ICSS (Itereted Cumulative Sums of Squares) permet de détecter des ruptures multiples dans des processus causaux. Le test est consistant en puissance et la comparaison avec des procédures existantes montre qu'il est plus puissant.

Le chapitre 4 étudie la détection des ruptures on-line dans la classe de modèle considéré aux chapitres 2 et 3. Une procédure basée sur la quasi-vraisemblance des observations a été développée. La procédure est consistante en puissance et le délai de détection est meilleur que celui des procédures existantes.

Le chapitre 5 est consacré aux applications aux débits du bassin versant de la Sanaga, les procédures décrites aux chapitres 2 et 3 ont été utilisées en appliquant un modèle ARMA sur les données désaisonnalisées et standardisées. Ces deux procédures ont détecté des ruptures qui sont "proches".

Mots-clés

Processus causaux, détection de rupture, modèle semi paramétrique, estimateur du maximum de quasi-vraisemblance, critère de vraisemblance pénalisé, convergence faible, débits hydrologiques.

Change-point detection in causal processes : Applications to the flows of the watershed of the Sanaga in Cameroon.

Abstract

This thesis focuses on the change-point detection in the causal processes with applications to the flows of the watershed of the Sanaga in Cameroon. We consider a class of semi-parametric models that contains the classical causal processes such as AR, ARCH, TARCH.

Chapter 1 is a summary of the works. It presents the model with examples and the main results obtained in chapters 2, 3 and 4.

Chapter 2 deals with the off-line multiple changes detection using a penalized likelihood criterion. The number of breaks, the dates of breaks and the parameters of model on each segment are unknown. They are estimated by maximizing a contrast based on the quasi-likelihood and penalized by the number of breaks. We suggest a possible choice of penalty parameter and show that the estimators of the parameters of model are consistent with optimal rates. For practical applications an adaptive estimator of the penalty parameter based on the slope heuristic is proposed. A dynamic programming algorithm is used to reduce the computational cost, it is now a $\mathcal{O}(n^2)$ complexity algorithm. Comparisons made with existing results show that our procedure is more stable and robust.

Chapter 3 is still the off-line multiple changes detection, but here a test procedure is used. We construct a new procedure that, combined with Itereted Cumulative Sums of Squares (ICSS) type algorithm, is able to detect multiple breaks in causal processes. The test is consistent in power and the comparison with existing procedures shows that it is more powerful.

In chapter 4, we study the on-line change detection in the class of semi-parametric model considered in chapters 2 and 3. A procedure based on the quasi-likelihood of the observations was developed. The procedure is consistent in power and the detection delay is better than existing ones.

Chapter 5 deals with applications to the flows of the watershed of the Sanaga. The procedures described in chapters 2 and 3 were used by applying an ARMA model after deseasonalization and standardization. Both procedures detected breaks which are "close".

Keywords

Causal processes, change-point detection, semi-parametric model, quasi-likelihood maximum estimator, penalized likelihood criterion, weak convergence, hydrological flows.

Table des matières

Chapitre 1

Synthèse des travaux

1.1 Introduction

Dans la vie de tous les jours, beaucoup de données sont souvent issues des modèles susceptibles de changer avec le temps. C'est le cas des données financières, hydrologiques... Plusieurs auteurs ont signalé le danger que l'on court si ces instants de changements ne sont pas étudiés et identifiés avant toute inférence statistique. Dès lors, beaucoup de recherches se sont penchées sur la détection de rupture.

Les premiers résultats sur la détection de ruptures datent de Page (1955), lorsqu'il utilisait les sommes partielles pour tester les changements dans la moyenne des observations indépendantes. Depuis lors, les recherches sur ce sujet ont considérablement avancé et ont permis de diviser ce problème en deux sous-classes : la détection off-line et la détection on-line. Dans le premier cas, on suppose que les données ont été entièrement observées, et on fait la détection de rupture dans ces données qui sont disponibles. Dans le cas on-line, on suppose que les données arrivent successivement dans le temps et que l'on essaye de détecter des ruptures éventuelles au fur et à mesure que les données arrivent. Le livre de Basseville et Nikiforov [11] propose une synthèse des méthodes statistiques utilisées pour ce problème.

De nos jours, des résultats importants ont été obtenus sur ce sujet (par exemple Inclan et Tiao [37], Davis *et al.* [24], Chu *et al.* [23], Berkes *et al.* [15], Davis *et al.* [25], Lee et Song [55], Aue *et al.* [3],...). Le cadre paramétrique reste le mieux étudié ; beaucoup de questions restent sans réponse dans les cas semi-paramétriques et non-paramétriques. Cette thèse apporte une nouvelle contribution sur ce sujet.

Ce chapitre reprend les principaux résultats obtenus dans [9], [40], [10] et détaillés dans les chapitres 2, 3 et 4. Après avoir présenté le modèle et les exemples, nous présentons une première procédure de détection off-line de rupture multiple basée sur un critère de vraisemblance pénalisée. Ensuite, nous développons une deuxième procédure de détection off-line de rupture multiple, basée sur une série de tests. Enfin, une nouvelle procédure de détection on-line de rupture est proposée.

Les procédures off-line développées aux chapitres 2 et 3 ont été appliquées au chapitre 5 sur les débits du bassin versant de la Sanaga. Les débits ont été standardisés et différenciés et un modèle ARMA a été appliqué. Les résultats montrent la présence des ruptures structurelles dans ces débits.

La Figure 1.1 présente la structure générale de la thèse, utile pour une lecture plus efficace de ce document.

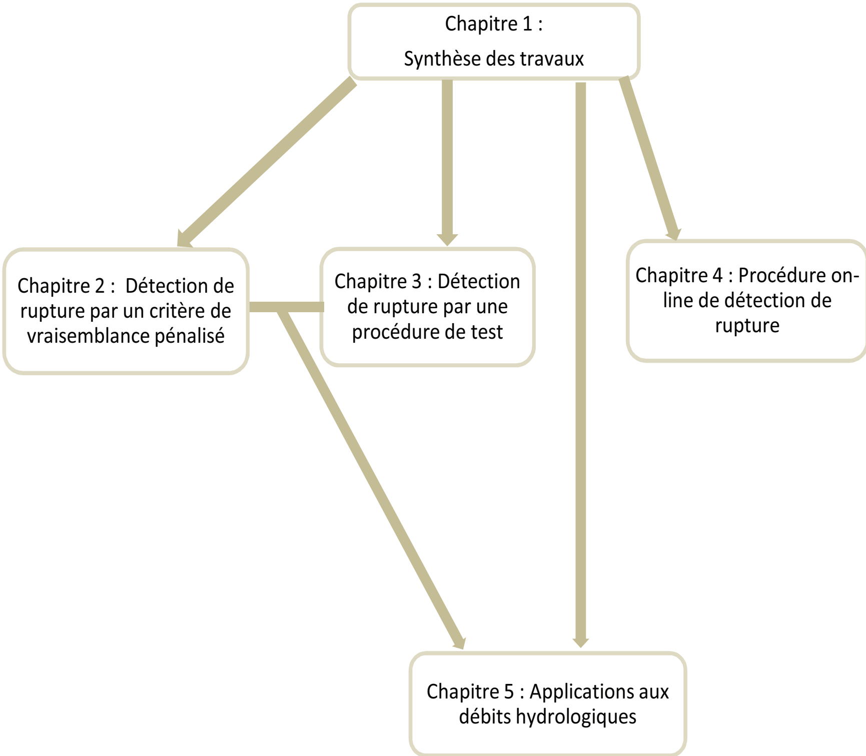

FIGURE 1.1 – Structure générale de la thèse et dependence entre les chapitres 1 à 5.

1.2 Présentation du modèle et les exemples

1.2.1 Présentation du modèle

Dans la pratique, beaucoup de phénomènes dynamiques évoluent en fonction de la loi de leur passé *i.e.* le comportement à un instant t ne dépend que de ce qui s'est passé jusqu'à l'instant $t-1$ et de l'"environnement" à l'instant t. Ces processus qui sont indépendants des événements futurs sont appelés processus causaux. La dynamique du processus admet généralement la forme

$$X_t = g(\xi_t, X_{t-1}, X_{t-2}, \cdots) \quad \text{pour tout } t \in \mathbb{Z} \tag{1.1}$$

où X_t représente la réalisation du processus à l'instant t, g une fonction mesurable et $(\xi_t)_{t\in\mathbb{Z}}$ une suite de variable aléatoire indépendante et identiquement distribuée (iid). L'existence et les propriétés d'un tel processus ont été étudiées par Doukhan et Wintenberger (2008). Nous étudierons une classe particulière de ce modèle.

Dans ce travail, on considère la classe de processus causaux définit pour tout $T \subset \mathbb{Z}$ par

Classe $\mathcal{M}_T(M_\theta, f_\theta)$ **:** *Le processus* $X = (X_t)_{t\in\mathbb{Z}}$ *appartient à* $\mathcal{M}_T(M_\theta, f_\theta)$ *s'il satisfait la relation :*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{pour tout } t \in T \tag{1.2}$$

où $\theta \in \Theta \subset \mathbb{R}^d$, M_θ et f_θ sont deux fonctions mesurables réelles telles que pour tout $(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ $M_\theta\big((x_i)_{i\in\mathbb{N}}\big) \neq 0$ et $(\xi_t)_{t\in\mathbb{Z}}$ une suite de variable aléatoire iid. Dans toute la suite, Θ est un compact fixé, dans lequel appartient le paramètre du model. On suppose que les formes des fonctions M_θ et f_θ sont connues et dépendent d'un paramètre θ qui est inconnu. Comme on va le voir ci-dessous, les processus classiques (tels que AR(∞), ARCH(∞), TARCH(∞)...) sont contenus dans la classe $\mathcal{M}_{\mathbb{Z}}(M_\theta, f_\theta)$.

Nous utiliserons dans la suite les normes suivantes :

1. $\|\cdot\|$ est la norme Euclidienne lorsqu'elle est appliquée à un vecteur ;
2. pour tout compact $\mathcal{K} \subseteq \mathbb{R}^d$ et pour toute fonction $g : \mathcal{K} \longrightarrow \mathbb{R}^{d'}$; $\|g\|_{\mathcal{K}} = \sup_{\theta\in\mathcal{K}}(\|g(\theta)\|)$;
3. pour tout $x = (x_1, \cdots, x_K) \in \mathbb{R}^K$, $\|x\|_m = \max\limits_{i=1,\cdots,K} |x_i|$;
4. si Y est un vecteur aléatoire admettant les moments jusqu'à l'ordre r, on pose $\|Y\|_r = (\mathbb{E}\|Y\|^r)^{1/r}$.

Pour garantir l'existence d'une unique solution stationnaire de la classe $\mathcal{M}_{\mathbb{Z}}(M_\theta, f_\theta)$, on fait les hypothèses suivantes pour $\Psi_\theta = M_\theta$, f_θ, $h_\theta = M_\theta^2$, $i = 0, 1, 2$ et $\mathcal{K} \subseteq \mathbb{R}^d$ un compact.

Hypothèse $\mathbf{A}_i(\Psi_\theta, \mathcal{K})$: *On suppose que* $\|\partial^i\Psi_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *et qu'il existe une suite de réels positifs* $(\alpha_k^{(i)}(\Psi_\theta, \mathcal{K}))_{k\geq1}$ *tel que* $\sum\limits_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K}) < \infty$ *et vérifie*

$$\Big\|\frac{\partial^i\Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i\Psi_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K})|x_k - y_k| \quad \textit{pour tout } x, y \in \mathbb{R}^{\mathbb{N}}.$$

Hypothèse $\mathbf{A}_i(h_\theta, \mathcal{K})$: *On suppose que* $f_\theta = 0$, $\|\partial^i h_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *et qu'il existe une suite de réels positifs* $(\alpha_k^{(i)}(h_\theta, \mathcal{K}))_{k\geq1}$ *tel que* $\sum\limits_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \mathcal{K}) < \infty$ *et vérifie*

$$\Big\|\frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \mathcal{K})|x_k^2 - y_k^2| \quad \textit{pour tout } x, y \in \mathbb{R}^{\mathbb{N}}.$$

Pour $i = 0, 1, 2$ et pour tout $\theta \in \Theta$, sous les hypothèses $A_i(f_\theta, \Theta)$ et $A_i(M_\theta, \Theta)$, posons :

$$\beta^{(i)}(\theta) := \sum_{k\geq1} \beta_k^{(i)}(\theta) \quad \text{where} \quad \beta_k^{(i)}(\theta) := \alpha_k^{(i)}(f_\theta, \{\theta\}) + (\mathbb{E}|\xi_0|^r)^{1/r}\alpha_k^{(i)}(M_\theta, \{\theta\}),$$

et sous l'hypothèse $A_i(h_\theta, \Theta)$

$$\tilde\beta^{(i)}(\theta) := \sum_{k\geq1} \tilde\beta_k^{(i)}(\theta) \quad \text{where} \quad \tilde\beta_k^{(i)}(\theta) := (\mathbb{E}|\xi_0|^r)^{2/r}\alpha_k^{(i)}(h_\theta, \{\theta\}).$$

La dépendance entre r et les coefficients $\beta^{(i)}(\theta)$ and $\tilde\beta^{(i)}(\theta)$ est omise dans le souci de simplifier les notations. Définissons l'ensemble

$$\Theta(r) := \Big\{\theta \in \Theta,\ A_0(f_\theta, \{\theta\}) \text{ et } A_0(M_\theta, \{\theta\}) \text{ satisfaites avec } \beta^{(0)}(\theta) < 1\Big\} \bigcup \Big\{\theta \in \Theta,\ f_\theta = 0 \text{ et } A_0(h_\theta, \{\theta\}) \text{ satisfaite avec } \tilde\beta^{(0)}(\theta) < 1\Big\}.$$

Proposition 1.1. *(Doukhan et Wintenberger [27]) Supposons $\theta \in \Theta(r)$ pour $r \geq 1$; alors il existe une unique solution causale $X = (X_t)_{t\in\mathbb{Z}} \in \mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$ stationnaire, ergodique et satistfaisant $\|X_0\|_r < \infty$.*

Soient $\theta \in \Theta(r)$ et $X = (X_t)_{t\in\mathbb{Z}}$ une solution stationnaire de la classe $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$. Pour la définition et l'étude de l'estimateur du maximum de quasi-vraisemblance, on fait les hypothèses supplémentaires suivantes.

Hypothèse D(Θ) **:** $\exists \underline{h} > 0$ tel que $\inf\limits_{\theta\in\Theta}(|h_\theta(x)|) \geq \underline{h}$ pour tout $x \in \mathbb{R}^{\mathbb{N}}$.

Hypothèse Id(Θ) : Pour tout $\theta, \theta' \in \Theta^2$,

$$\Big(f_\theta(X_0, X_{-1}, \cdots) = f_{\theta'}(X_0, X_{-1}, \cdots) \text{ et } h_\theta(X_0, X_{-1}, \cdots) = h_{\theta'}(X_0, X_{-1}, \cdots) \text{ a.s.}\Big) \Rightarrow \ \theta = \theta'.$$

Hypothèse Var(Θ) : pour tout $\theta \in \Theta$, une des familles $\big(\dfrac{\partial f_\theta}{\partial \theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ ou $\big(\dfrac{\partial h_\theta}{\partial \theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ est p.s. linéairement indépendant.

1.2.2 Examples

1. Modèles AR(∞)
Considérons le processus AR(∞) défini par :

$$X_t = \sum_{k\geq 1} \phi_k(\theta_0^*) X_{t-k} + \xi_t \ , \ t \in \mathbb{Z}$$

avec $\theta_0^* \in \Theta$, où Θ est un compact de $\mathbb{R}^d$ tel que $\sum_{k\geq 1} \|\phi_k(\theta)\|_\Theta < 1$. Le processus appartient à la classe $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ où $f_\theta(x_1, \cdots) = \sum_{k\geq 1} \phi_k(\theta) x_k$ et $M_\theta \equiv 1$ pour tout $\theta \in \Theta$. Anisi, les hypothèses D(Θ) et $A_0(f_\theta, \Theta)$ sont vérifiées avec $\underline{h} = 1$ et $\alpha_k^{(0)}(f_\theta, \Theta) = \|\phi_k(\theta)\|_\Theta$. De plus, si ξ_0 est une variable aléatoire non dégénérée (*ie* n'est pas égale à une constante), alors Id(Θ) et Var(Θ) sont vraies. Enfin, pour tout $r \geq 1$ tel que $\mathbb{E}|\xi_0|^r < \infty$, $\Theta(r) = \Theta = \{\theta \in \mathbb{R}^d ; \ \sum_{k\geq 1} |\phi_k(\theta)| < 1\}$.

2. Modèles ARCH(∞)
Considérons le processus $ARCH(\infty)$ défini par :

$$X_t = \sqrt{\psi_0(\theta_0^*) + \sum_{k\geq 1} \psi_k(\theta_0^*) X_{t-k}^2}\, \xi_t \ ,$$

avec $\theta_0^* \in \Theta$, où Θ est un compact de $\mathbb{R}^d$ tel que $\sum_{k\geq 1} \|\psi_k(\theta)\|_\Theta < 1$ $(\psi_k(\theta))_{k\geq 1}$ étant une suite de réels positifs. On suppose $\|\psi_0(\theta)\|_\Theta > 0$ et $\mathbb{E}(\xi_0^2) = 1$. Le processus appartient à la classe $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ où $M_\theta(x_1, \cdots) = \sqrt{\psi_0(\theta) + \sum_{k\geq 1} \psi_k(\theta) x_k}$ et $f_\theta \equiv 0 \ \forall \theta \in \Theta$. Supposons, $\inf_{\theta\in\Theta} \psi_0(\theta) > 0$. Alors, les hypothèses D(Θ) et $A_0(h_\theta, \Theta)$ sont vérifiées avec $\underline{h} = \inf_{\theta\in\Theta} \psi_0(\theta)$ et $\alpha_k^{(0)}(h_\theta, \Theta) = \|\psi_k(\theta)\|_\Theta$. De plus, si ξ_0 est une variable aléatoire non dégénérée, alors Id(Θ) et Var(Θ) sont vraies. Pour tout $r > 1$, on a $\Theta(r) = \{\theta \in \Theta \ ; \ (\mathbb{E}|\xi_0|^r)^{2/r} \sum_{k\geq 1} \|\psi_k(\theta)\|_\Theta < 1\}$.

3. TARCH(∞) model.
On définit le processus $TARCH(\infty)$ par :

$$X_t = \sigma_t\, \xi_t \ , \ \sigma_t = b_0(\theta_0^*) + \sum_{k\geq 1} \Big(b_k^+(\theta_0^*)\max(X_{t-k}, 0) - b_k^-(\theta_0^*)\min(X_{t-k}, 0)\Big), \ t \in \mathbb{Z}$$

avec $\theta_0^* \in \Theta$, où Θ est un compact de $\mathbb{R}^d$ stisfaisant $\sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < \infty$. On reconnait la classe $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ avec $f_\theta \equiv 0$ et $M_\theta(x_1, \cdots) = b_0(\theta) + \sum_{k\geq 1} \Big(b_k^+(\theta)\max(x_k, 0) - b_k^-(\theta)\min(x_k, 0) \quad \forall \theta \in \Theta$. L'hypothèse $(A_0(M_\theta, \Theta))$ est vérifiée avec $\alpha_k^{(0)}(M_\theta, \Theta) = \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta)$. On suppose que pour tout $\theta \in \Theta$, $b_0(\theta)$, $b_k^+(\theta)$, $b_k^-(\theta)$ sont des réels positifs et que $\inf_{\theta\in\Theta}(b_0(\theta)) > 0$. Alors, D($\Theta$) est vérifiée avec $\underline{h} = \inf_{\theta\in\Theta}(b_0(\theta))$. Si ξ_0 est une variable aléatoire non dégénérée, alors Id(Θ) et Var(Θ) sont vraies. Pour tout $r > 1$, on a

$$\Theta(r) = \big\{\theta \in \Theta\ ;\ (\mathbb{E}|\xi_0|^r)^{1/r} \sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < 1\big\}.$$

1.2.3 Estimateur du maximum de quasi-vraisemblance

On considère un processus $X = (X_t)_{t\in\mathbb{Z}}$ appartenant à la classe $\mathcal{M}_{\mathbb{Z}}(f_{\theta_0^*}, M_{\theta_0^*})$ dépendant d'un paramètre $\theta_0^* \in \Theta \subset \mathbb{R}^d$ que l'on souhaite estimer à partir des observations $X_1, \ldots, X_n$.

Supposons dans un premier temps que $(\xi_t)_{t\in\mathbb{Z}}$ est un processus gaussien. Par définition, la moyenne et la variance conditionnées par le passé sont respectivement $f_{\theta_0^*}(X_{s-1}, \ldots)$ et $h_{\theta_0^*}(X_{s-1}, \ldots)$. Ainsi, on a

$$X_t \mid (X_{t-j})_{j\in\mathbb{N}^*} \sim \mathcal{N}\Big(f_{\theta_0^*}(X_{t-1}, \ldots), h_{\theta_0^*}(X_{s-1}, \ldots)\Big).$$

En posant $f_\theta^s = f_\theta(X_{s-1}, X_{s-2}\ldots)$, $M_\theta^s = M_\theta(X_{s-1}, X_{s-2}\ldots)$ et $h_\theta^s = M_\theta^{s\,2}$ on déduit la fonction de quasi-vraisemblance calculée sur un segment $T \subset \mathbb{Z}$

$$L_n(T, \theta) := -\frac{1}{2}\sum_{s\in T} q_s(\theta) \ \text{ with } \ q_s(\theta) := \frac{(X_s - f_\theta^s)^2}{h_\theta^s} + \log\left(h_\theta^s\right). \tag{1.3}$$

Par convention, on posera $L_n(\emptyset, \theta_k) := 0$.

Ayant observé $X_1, \ldots, X_n$, cette vraisemblance n'est pas calculable car elle dépend des valeurs $(X_{-j})_{j\in\mathbb{N}}$ qui ne sont pas observées. Ainsi on définit la log-vraisemblance approchée par :

$$\widehat{L}_n(T, \theta) := -\frac{1}{2}\sum_{s\in T} \widehat{q}_s(\theta) \quad \text{where} \quad \widehat{q}_s(\theta) := \frac{(X_s - \widehat{f}_\theta^s)^2}{\widehat{h}_\theta^s} + \log\big(\widehat{h}_\theta^s\big). \tag{1.4}$$

avec $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \ldots, X_1, 0, 0, ...)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \ldots, X_1, 0, 0, ...)$ et $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$. On définit alors l'estimateur du maximum de quasi-vraisemblance du paramètre θ_0^* calculé sur T par

$$\widehat{\theta}_n(T) := \underset{\theta\in\Theta}{\operatorname{argmax}}(\widehat{L}_n(T, \theta)).$$

1.3 Détection off-line de rupture par un critère de vraisemblance pénalisé

1.3.1 Présentation du problème

On considère un processus $X = (X_t)_{t\in\mathbb{Z}}$ appartenant à la classe de modèle définie ci-dessus dont on a fait des observations $X_1, \ldots, X_n$. On suppose que X dépend d'un paramètre θ_0^* susceptible de changer K^*-1 fois durant les observations. Plus précisément, cela signifie

$$X \in \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*}) \quad \text{pour } j = 1, \dots, K^* \tag{1.5}$$

où

- $K^* \in \mathbb{N}^*$, représente le vrai nombre de segments ;
- $T_j^* = \{t_{j-1}^*+1, t_{j-1}^*+2, \dots, t_j^*\}$ avec $0 < t_1^* < \dots < t_{K^*-1}^* < n$, $t_j^* \in \mathbb{N}$ et par convention $t_0^* = -\infty$ and $t_{K^*}^* = \infty$;
- $\theta_j^* \in \Theta \subset \mathbb{R}^d$ pour $j = 1, \dots, K^*$ représente le paramètre du modèle sur le segment T_j^*.

Dans ce modèle, K^*, $(t_1^*, \dots, t_{K^*-1}^*)$ et $(\theta_1^*, \theta_2^*, \cdots, \theta_{K^*}^*)$ sont inconnus. L'objectif de cette partie est de construire des estimateurs consistants de ces paramètres. Nous utiliserons les notations suivantes dans la suite.

Notation .
- *Pour $K \geq 2$, $\mathcal{F}_K = \{\underline{t} = (t_1, \dots t_{K-1})\ ;\ 0 < t_1 < \dots < t_{K-1} < n\}$. En particulier, $\underline{t}^* = (t_1^*, \dots, t_{K^*-1}^*) \in \mathcal{F}_{K^*}$ représente le vecteur des vrais instants de rupture ;*
- *Pour $K \in \mathbb{N}^*$ et $\underline{t} \in \mathcal{F}_K$, $T_k = \{t \in \mathbb{Z},\ t_{k-1} < t \leq t_k\}$ et $n_k = Card(T_k)$ avec $1 \leq k \leq K$. En particulier ; $T_j^* = \{t \in \mathbb{Z},\ t_{j-1}^* < t \leq t_j^*\}$ et $n_j^* = Card(T_j^*)$ pour $1 \leq j \leq K^*$. Pour $1 \leq k \leq K$ et $1 \leq j \leq K^*$, posons $n_{kj} = Card(T_j^* \cap T_k)$;*

La proposition suivante garantit l'existence d'une solution du problème 1.5.

Proposition 1.2. *En référence au problème* (1.5) *; supposons qu'il existe $r \geq 1$ tel que $\theta_j^* \in \Theta(r)$ pour tout $j = 1, \dots, K^*$. Alors*

(i) il existe un processus causal $X = (X_t)_{t\in\mathbb{Z}}$ solution du modèle (1.5) *tel que $\|X_t\|_r < \infty$ pour tout $t \in \mathbb{Z}$.*

(ii) il existe une constante $C > 0$ telle que pour tout $t \in \mathbb{Z}$, $\|X_t\|_r \leq C$.

1.3.2 Définition des estimateurs

Considérons un processus $X = (X_t)_{t\in\mathbb{Z}}$ solution du problème (1.5). On suppose qu'on a observé une trajectoire $(X_1, \cdots, X_n)$ de X. Pour tout $K \geq 1$, $\underline{t} \in \mathcal{F}_K$ et $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_K) \in \Theta(r)^K$, définissons le critère non-pénalisé $QLIK$ par :

$$(QLIK) \quad \widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2\sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k).$$

et le critère pénalisé *penQLIK* par

$$(penQLIK) \quad \widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n K \tag{1.6}$$

où $\kappa_n < n$ représente le paramètre de pénalité. On suppose qu'on connait une borne supérieure K_{max} du nombre de ruptures. L'estimateur des paramètres du modèle est défini par :

$$(\widehat{K}_n, \underline{\widehat{t}}_n, \underline{\widehat{\theta}}_n) \in \underset{1\leq K\leq K_{\max}}{\text{Argmin}} \ \underset{(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta(r)^K}{\text{Argmin}} (\widetilde{J}_n(K, \underline{t}, \underline{\theta})) \quad \text{et} \quad \underline{\widehat{\tau}}_n = \frac{\underline{\widehat{t}}_n}{n}. \tag{1.7}$$

Comme on peut le remarquer, l'estimateur ci-dessus dépend fortement du paramètre de régularité κ_n. Ce paramètre doit donc être convenablement choisi pour garantir la consistance asymptotique de l'estimateur. On fait l'hypothèse suivante

Hypothèse $\mathbf{H}_i$ $(i = 0, 1, 2)$: *Pour $0 \le p \le i$, $A_p(f_\theta, \Theta)$, $A_p(M_\theta, \Theta)$ (ou respectivement $A_p(h_\theta, \Theta)$) sont satisfaites et pour tout $j = 1, \ldots, K^*$ il existe $r \ge 1$ tel que $\theta_j^* \in \Theta(r)$. En posant*

$$c^* = \min_{j=1,\cdots,K^*} \big(- \log(\beta^{(0)}(\theta_j^*))/8\big) \quad \textit{or resp.} \quad \min_{j=1,\cdots,K^*} \big(- \log(\tilde{\beta}^{(0)}(\theta_j^*))/8\big)$$

le paramètre (κ_n) utilisé dans (2.4) vérifie $\kappa_n \wedge n\,\kappa_n^{-1} \to \infty$ avec $n \to \infty$ et pour tout $j = 1, \cdots K^*$:

$$\sum_{k \ge 2} \kappa_k^{-(r/4\wedge 1)} \Big(\sum_{\ell \ge k c^*/\log(k)} \beta_\ell^{(p)}(\theta_j^*) \Big)^{(r/4\wedge 1)} < \infty \ \textit{ou resp.} \ \sum_{k \ge 2} \kappa_k^{-(r/4\wedge 1)} \Big(\sum_{\ell \ge k c^*/\log(k)} \tilde{\beta}_\ell^{(p)}(\theta_j^*) \Big)^{(r/4\wedge 1)} < \infty \quad (1.8)$$

Les conditions H_i paraissent très complexes, pourtant elles sont vérifiées par l'ensemble des modèles classiques cités précédemment. Donnons deux cas classiques pour lesquels elles sont vérifiées

(1) cas géométrique : si $\alpha_\ell^{(i)}(f_\theta, \Theta(r)) + \alpha_\ell^{(i)}(M_\theta, \Theta(r)) + \alpha_\ell^{(i)}(h_\theta, \Theta(r)) = O(a^\ell)$ avec $0 \le a < 1$, alors tout de (κ_n) tel que $\kappa_n \to \infty$ et $\kappa_n = o(n)$, satisfait (1.8).

(2) Cas Riemanien : si $\alpha_\ell^{(i)}(f_\theta, \Theta(r)) + \alpha_\ell^{(i)}(M_\theta, \Theta(r)) + \alpha_\ell^{(i)}(h_\theta, \Theta(r)) = O(\ell^{-\gamma})$ avec $\gamma > 1$,
- si $\gamma > 1 + (1 \vee 4r^{-1})$, alors tout choix de (κ_n) tel que $\kappa_n \to \infty$ et $\kappa_n = o(n)$ satisfait (1.8).
- si $(1 \vee 4r^{-1}) < \gamma \le 1 + (1 \vee 4r^{-1})$, alors tout choix de (κ_n) tel que $O(\kappa_n) = n^{1-\gamma+(1\vee 4r^{-1})}(\log n)^\delta$ avec $\delta > \gamma - 1 + (1 \vee 4r^{-1})$ et $\kappa_n = o(n)$ peut être choisi.

.

1.3.3 Résultats asymptotiques

Pour l'étude asymptotique des estimateurs, on fait les hypothèses classiques suivantes
Hypothèse B : $\displaystyle \mathop{min}_{j=1,\cdots,K^*-1} \|\theta_{j+1}^* - \theta_j^*\| > 0$.

Hypothèse C : *il existe $\underline{\tau}^* = (\tau_1^*, \ldots, \tau_{K^*-1}^*)$ avec $0 < \tau_1^* < \ldots < \tau_{K^*-1}^* < 1$ tel que pour $j = 1, \cdots, K^*$, $t_j^* = [n\tau_j^*]$ (où $[x]$ désigne la partie entière de x). $\underline{\tau}^*$ est appélé le vecteur des ruptures.*

Théorème 1.3. *On suppose que les hypothèses $D(\Theta(r))$, $Id(\Theta(r))$, B, C et H_0 sont satisfaites avec $r \ge 2$. Si $K_{\max} \ge K^*$ alors :*

$$(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (K^*, \underline{\tau}^*, \underline{\theta}^*).$$

Le Théorème 1.4 et le Théorème 1.5 ci-dessous donnent respectivement la vitesse de convergence de l'estimateur des instants de rupture $\widehat{\underline{\tau}}_n$ et la normalité asymptotique de l'estimateur $\widehat{\underline{\theta}}_n$.

Théorème 1.4. *On suppose que les hypothèses $D(\Theta(r))$, $Id(\Theta(r))$, B, C and H_2 sont satisfaites avec $r \ge 4$. Si $K_{\max} \ge K^*$ alors*

$$\lim_{\delta \to \infty} \lim_{n \to \infty} \mathbb{P}(\|\widehat{\underline{t}}_n - \underline{t}^*\|_m > \delta) = 0. \qquad (1.9)$$

Il découle du Théorème 1.4 que si $(w_n)_n$ est une suite tendant vers l'infini, alors $w_n^{-1}\,\|\widehat{\underline{t}}_n - \underline{t}^*\|_m \xrightarrow{P} 0$. Ce qui montre que la vitesse de convergence de $\widehat{\underline{\tau}}_n$ vers $\underline{\tau}^*$ est de l'ordre de $\frac{1}{n}$.

Pour la normalité asymptotique de l'estimateur $\widehat{\underline{\theta}}_n$, si $\widehat{K}_n < K^*$, poser $\widehat{T}_j = \widehat{T}_{\widehat{K}_n}$ pour $j \in \{\widehat{K}_n, \ldots, K^*\}$. On a le théorème

Théorème 1.5. *On suppose que les hypothèses $D(\Theta(r))$, $Id(\Theta(r))$, B, C and H_2 sont satisfaites avec $r \geq 4$ et $\kappa_n = O(\sqrt{n})$. Si $\theta_j^* \in \overset{\circ}{\Theta}(r)$ pour tout $j = 1, \cdots, K^*$, alors*

$$\sqrt{n_j^*}\,(\widehat{\theta}_n(\widehat{T}_j) - \theta_j^*) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, F(\theta_j^*)^{-1}G(\theta_j^*)F(\theta_j^*)^{-1}), \tag{1.10}$$

les matrices F et G sont données par

$$(F(\theta_j^*))_{k,l} = I\!E\Big(\frac{\partial^2 q_{0,j}(\theta_j^*)}{\partial\theta_k\partial\theta_l}\Big) \text{ et } (G(\theta_j^*))_{k,l} = I\!E\Big(\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_k}\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_l}\Big). \tag{1.11}$$

où pour tou $t \in \mathbb{Z}$, la fonction $\theta \mapsto q_{t,j}(\theta)$ est définie comme en (1.3) mais calculée à partir de la solution stationnaire de la classe $\mathcal{M}_{\mathbb{Z}}(f_{\theta_j^*}, M_{\theta_j^*})$.

1.3.4 Simulations

Nous allons faire des simulations sur un modèle AR(1). La procédure est implémentée à partir du logiciel R. La programmation dynamique (voir [39]) est utilisée pour réduire la complexité numérique de la procédure, qui est désormais de l'ordre de $\mathcal{O}(n^2)$.

Nous avons donné ci-dessus les choix possibles du paramètre κ_n permettant d'obtenir la consistance de l'estimateur $\widehat{K}_n$. Mais, ces choix sont asymptotiques, pour des applications pratiques (avec généralement $n \leq 2000$), nous proposons un choix adaptatif du paramètre κ_n. Ce choix est basé sur le principe de l'heuristique de la pente (voir Baudry *et al.*, 2010 [12]).

Heuristique de la pente

Cette heuristique fait l'hypothèse que le critère $-QLIK$ est une fonction linéaire de K à partir d'un certain rang. Plus précisément, elle consiste à :

- représenter la courbe $(K, -\min_{\underline{t},\underline{\theta}} QLIK(K))_{1\leq K\leq K_{max}}$;
- calculer la pente de la partie linéaire de cette courbe, que l'on note $\hat{\kappa}_n/2$;
- utiliser $\kappa_n = \hat{\kappa}_n$ comme paramètre de pénalité dans le critère pénalisé.

Quelques résultats de simulations pour le modèle AR(1)

On considère le problème (1.5) pour un processus AR(1) :

$$X_t = \theta_j^*\, X_{t-1} + \xi_t \qquad \forall t \in T_j^*, \quad \forall j \in \{1, \cdots, K^*\}.$$

Pour $n = 500$ et $n = 1000$, on simule une trajectoire $(X_1, \cdots, X_n)$ avec deux ruptures ($K^* = 3$) : le paramètre $\theta^*(1) = 0.7$ change en $\theta^*(2) = 0.9$ à l'instant $t_1^* = 0.3n$ qui change aussi en $\theta^*(3) = 0.6$ à l'instant $t_2^* = 0.7n$ (ainsi $(\tau_1^*, \tau_2^*) = (0.3, 0.7)$). La Figure

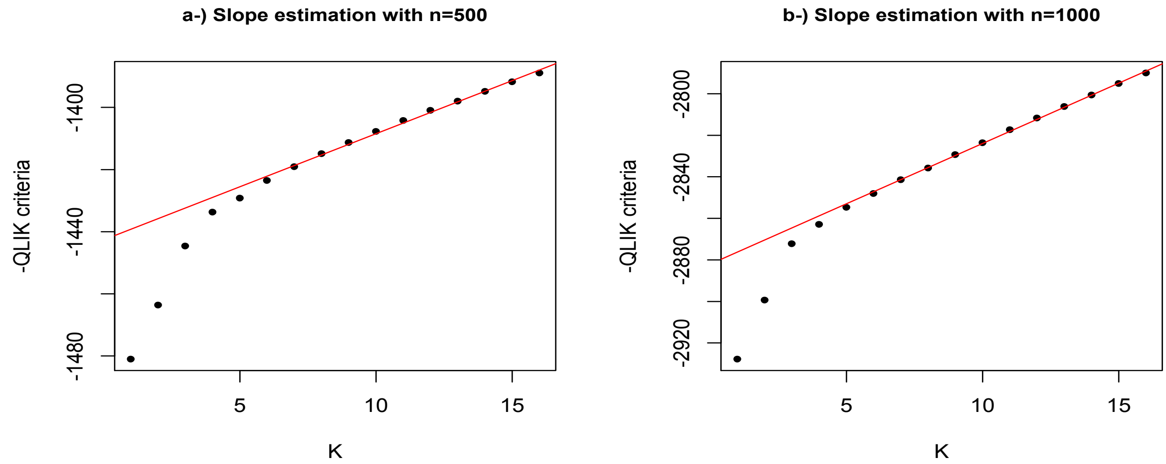

Figure 1.2 – La courbe de $-\min_{t,\theta} QLIK$ pour $1 \leq K \leq K_{max} = 16$. La pente de la partie linéaire est $\hat{\kappa}_n/2 = 3.47$ pour $n = 500$ et $\hat{\kappa}_n/2 = 5.75$ pour $n = 1000$.

1.2 représente la pente de la partie linéaire du critère $-QLIK$ (minimisé en $(\underline{t}, \underline{\theta})$) pour $n = 500$ et $n = 1000$. En se reférant à la Figure 1.2, on obtient $\hat{\kappa}_n \approx 6.94$ pour $n = 500$ et $\hat{\kappa}_n \approx 11.50$ pour $n = 1000$. On peut maintenant minimiser le critère pénalisé pour estimer le nombre de ruptures. La Figure 1.3 représente les points $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ pour $1 \leq K \leq K_{max} = 10$.

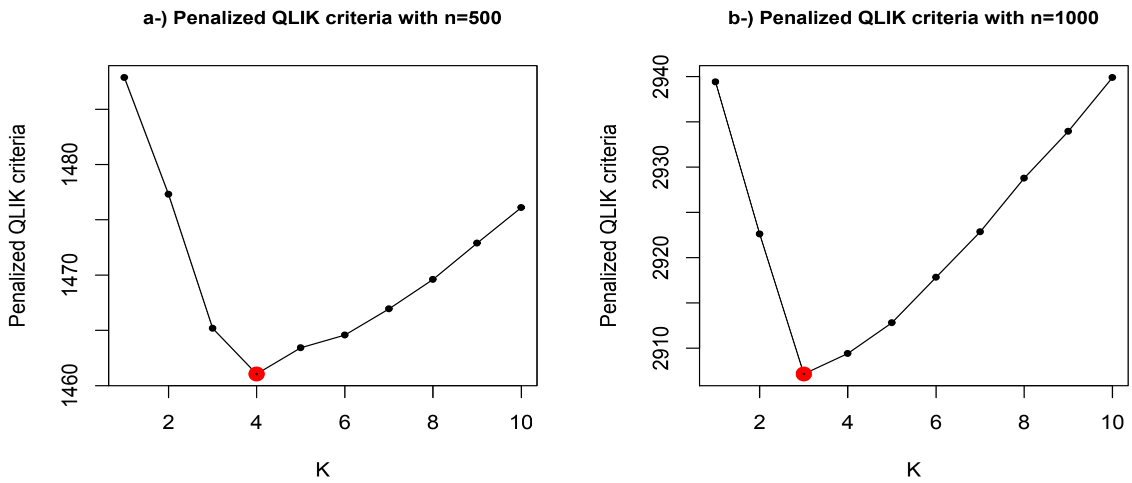

Figure 1.3 – Courbe du critère pénalisé.

On peut facilement lire sur la Figure 1.3 $\widehat{K}_n = 4$ pour $n = 500$ et $\widehat{K}_n = 3$ pour $n = 1000$ (le nombre de ruptures estimé étant $\widehat{K}_n$ -1). De plus, les instants de ruptures estimés sont $\widehat{\underline{t}}_n = (137, 253, 345)$ ($\underline{t}^* = (150, 350)$) pour $n = 500$ et $\widehat{\underline{t}}_n = (323, 675)$ ($\underline{t}^* = (300, 700)$) pour $n = 1000$. La Figure 1.4 montre les instants de ruptures estimés et les vrais instants de rupture dans le processus.

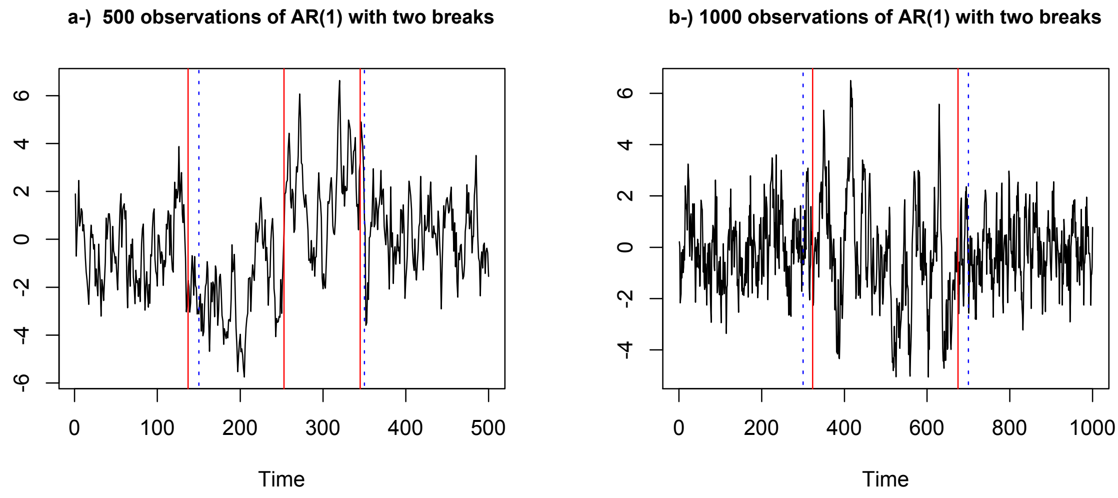

FIGURE 1.4 – Un processus AR(1) avec deux ruptures. Le paramètre $\theta^*(1) = 0.7$ change en $\theta^*(2) = 0.9$ à l'instant $t_1^* = 0.3n$ qui change à son tour en $\theta^*(3) = 0.6$ à l'instant $t_2^* = 0.7n$ ($K^* = 3$). Les traits forts représentent les instants de ruptures estimés et les traits en pointillés représentent les vrais instants de ruptures.

1.4 Détection off-line de rupture par une procédure de test

1.4.1 Présentation du problème

On considère toujours $X = (X_t)_{t\in\mathbb{Z}}$ appartenant à la famille de modèle $\mathcal{M}_T(f_\theta, M_\theta)$ (définit en (1.2)) avec $\theta \in \Theta$. On suppose qu'on a observé une trajectoire $(X_1, \dots, X_n)$ de X. On repose le problème (1.5) sous forme d'un test d'hypothèse comme suit :

$\mathbf{H}_0$: il existe $\theta_0 \in \Theta$ tel que $(X_1, \cdots, X_n)$ appartient à la classe $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0}, f_{\theta_0})$;
$\mathbf{H}_1$: il existe $K \geq 2$, $\theta_1^*, \cdots, \theta_K^* \in \Theta$ avec $\theta_j^* \neq \theta_{j+1}^*$, tel que $(X_1, \cdots, X_n)$ appartient à $\bigcap_{j=1}^{K} \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*})$ où $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \cdots, t_j^*\}$ avec $0 = t_0^* < t_1^* < \cdots < t_{K-1}^* < t_K^* = n$.

L'objectif est de développer une procédure convenable permettant de tester $\mathbf{H}_0$ contre $\mathbf{H}_1$.

1.4.2 Définition de la statistique de test

On suppose qu'une trajectoire $(X_1, \dots, X_n)$ du processus a été observée. Rappelons (voir aussi 1.4) que pour tout $T \subset \{1, \cdots, n\}$ la quasi-vraisemblance approchée calculée sur le segment T est donnée par :

$$\widehat{L}_n(T, \theta) := -\frac{1}{2} \sum_{t\in T} \widehat{q}_t(\theta) \quad \text{avec} \quad \widehat{q}_t(\theta) := \frac{(X_t - \widehat{f}_\theta^t)^2}{\widehat{h}_\theta^t} + \log\left(\widehat{h}_\theta^t\right)$$

où $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \dots, X_1, 0, 0, \cdots)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \dots, X_1, 0, 0, \cdots)$ et $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$. On définit aussi l'estimateur du maximum de quasi-vraisemblance calculé sur T par $\widehat{\theta}_n(T) := \underset{\theta\in\Theta}{\operatorname{argmax}}(\widehat{L}_n(T, \theta))$.

Pour tout $T \subset \{1, \cdots, n\}$, définissons les matrices

$$\widehat{G}_n(T) := \frac{1}{Card(T)} \sum_{t \in T} \Big(\frac{\partial \widehat{q}_t(\widehat{\theta}_n(T))}{\partial \theta} \Big) \Big(\frac{\partial \widehat{q}_t(\widehat{\theta}_n(T))}{\partial \theta} \Big)' \tag{1.12}$$

et

$$\widehat{F}_n(T) := -\frac{2}{Card(T)} \Big(\frac{\partial^2 \widehat{L}_n(T, \widehat{\theta}_n(T))}{\partial \theta \partial \theta'} \Big) = \frac{1}{Card(T)} \sum_{t \in T} \frac{\partial^2 \widehat{q}_t(\widehat{\theta}_n(T))}{\partial \theta \partial \theta'}. \tag{1.13}$$

Pour $k = 1, \cdots, n-1$, posons $T_k = \{1, \cdots, k\}$, $\overline{T}_k = \{k+1, \cdots, n\}$ et définissons

$$\widehat{\Sigma}_{n,k} := \frac{k}{n} \widehat{F}_n(T_k) \widehat{G}_n(T_k)^{-1} \widehat{F}_n(T_k) \mathbb{1}_{\det(\widehat{G}_n(T_k)) \neq 0} + \frac{n-k}{n} \widehat{F}_n(\overline{T}_k) \widehat{G}_n(\overline{T}_k)^{-1} \widehat{F}_n(\overline{T}_k) \mathbb{1}_{\det(\widehat{G}_n(\overline{T}_k)) \neq 0}.$$

Soit $(v_n)_{n \in \mathbb{N}}$ une suite numérique satisfaisant $v_n \to \infty$ et $v_n/n \to 0$ (quand $n \to \infty$). Posons en fin $\Pi_n = [v_n, n - v_n] \cap \mathbb{N}$ et définissons les statistiques :

$$\widehat{Q}_n^{(1)} := \max_{k \in \Pi_n} \widehat{Q}_{n,k}^{(1)} \quad \text{avec} \quad \widehat{Q}_{n,k}^{(1)} := \frac{k^2}{n} \big(\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n) \big)' \widehat{\Sigma}_{n,k} \big(\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n) \big);$$

$$\widehat{Q}_n^{(2)} := \max_{k \in \Pi_n} \widehat{Q}_{n,k}^{(2)} \quad \text{avec} \quad \widehat{Q}_{n,k}^{(2)} := \frac{(n-k)^2}{n} \big(\widehat{\theta}_n(\overline{T}_k) - \widehat{\theta}_n(T_n) \big)' \widehat{\Sigma}_{n,k} \big(\widehat{\theta}_n(\overline{T}_k) - \widehat{\theta}_n(T_n) \big).$$

La statistique de test est définie par

$$\widehat{Q}_n := \max\big(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}\big) = \max_{k \in \Pi_n} \Big(\max\big(\widehat{Q}_{n,k}^{(1)}, \widehat{Q}_{n,k}^{(2)}\big) \Big)$$

Pour des applications pratiques, nous recommandons d'utiliser $v_n = [(\log n)^2]$ pour des processus linéaires et $v_n = [(\log n)^\delta]$ (avec $5/2 \le \delta < 3$) pour des modèles de type GARCH et TARCH.

1.4.3 Résultats asymptotiques

Pour l'étude asymptotique, on fait l'hypothèse suivante :

Hypothèse K$(f_\theta, M_\theta, \Theta)$: pour i= 0, 1, 2, $\mathbf{A}_i(f_\theta, \Theta)$ and $\mathbf{A}_i(M_\theta, \Theta)$ (ou $\mathbf{A}_i(h_\theta, \Theta)$) sont vraies et il existe $\ell > 2$ tel que $\alpha_j^{(i)}(f_\theta, \Theta) + \alpha_j^{(i)}(M_\theta, \Theta) + \alpha_j^{(i)}(h_\theta, \Theta) = \mathcal{O}(j^{-\ell})$, for i= 0, 1.

Théorème 1.6. *Supposons que les hypothèses* $\boldsymbol{D}(\Theta)$, ***Id***(Θ), ***Var*** *et* ***K***$(f_\theta, M_\theta, \Theta)$ *sont satisfaites. Sous* H_0, *si* $\theta_0 \in \overset{\circ}{\Theta}(4)$, *alors pour* $j = 1, 2$,

$$\widehat{Q}_n^{(j)} \xrightarrow[n \to \infty]{\mathcal{D}} \sup_{0 \le \tau \le 1} \|W_d(\tau)\|^2$$

où W_d *est un pont brownien de dimension* d.

Pour tout $\alpha \in (0, 1)$, désignons par C_α le quantile d'ordre $(1 - \alpha/2)$ de la loi de $\sup\limits_{0 \le \tau \le 1} \|W_d(\tau)\|^2$. Alors, le corollaire suivant s'ensuit immédiatement.

Corollaire 1.7. *Sous les hypothèses du Théorème 1.6 :*

$$\forall \alpha \in (0,1) \quad \limsup_{n \to \infty} P\big(\widehat{Q}_n > C_\alpha\big) \le \alpha.$$

Les quantiles de la loi de $\sup_{0\leq\tau\leq1} \|W_d(\tau)\|^2$ sont connus, ou peuvent être facilement calculés par une procédure de Monte Carlo. D'après le Théorème 1.6 et le Corollaire 1.7, une grande valeur de $\widehat{Q}_n$ indiquerait la présence des ruptures dans les observations. A un seuil α donné, on prend comme région critique l'ensemble $(\widehat{Q}_n > C_\alpha)$.
La Figure 1.5 est une illustration pour un processus AR(1).

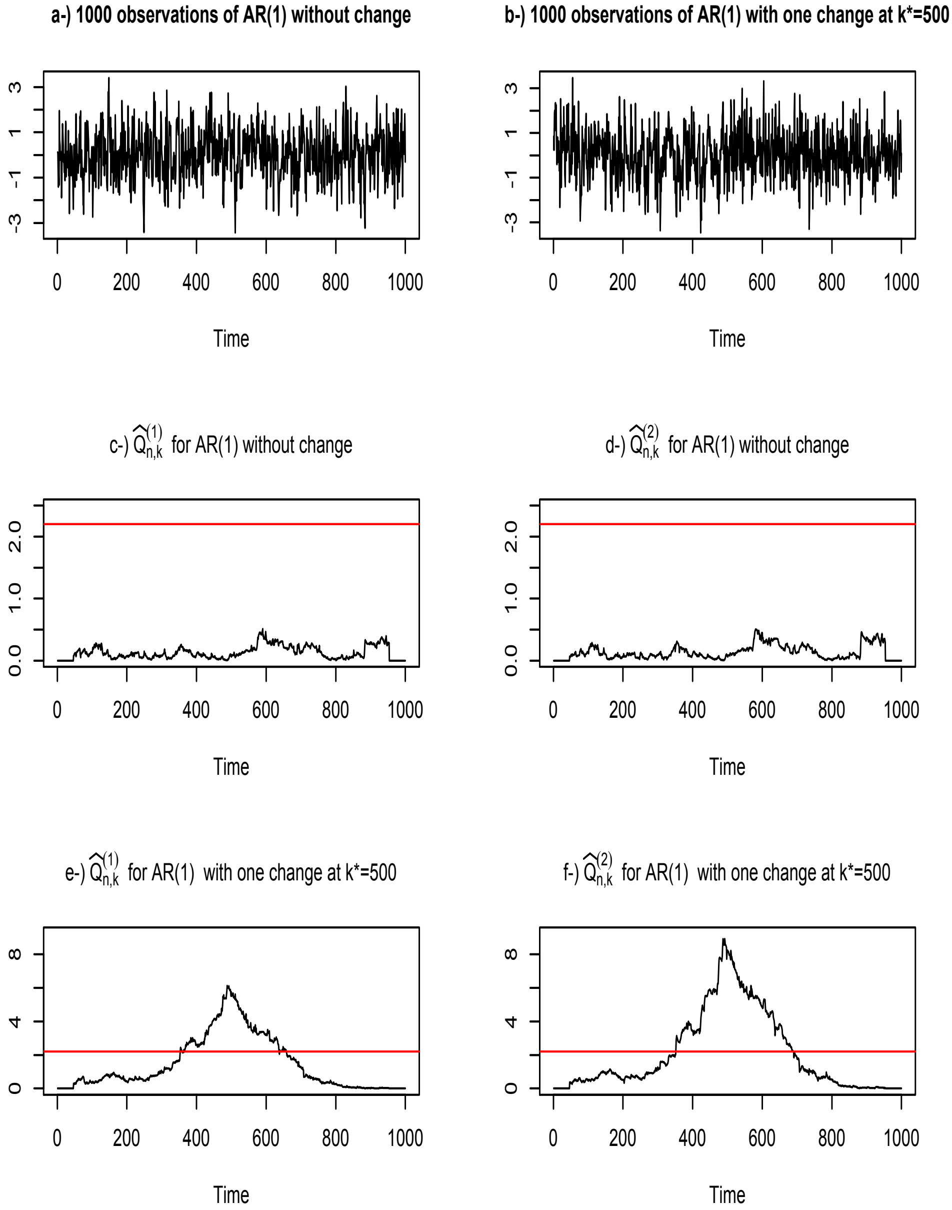

FIGURE 1.5 – 1000 observation d'un processus AR(1) et les statistiques $\widehat{Q}^{(1)}_{n,k}$ et $\widehat{Q}^{(2)}_{n,k}$ correspondantes avec $v_n = [(\log n)^2]$. a-) un processus AR(1) sans rupture, avec pour paramètre $\phi_1 = 0.4$. b-) un processus AR(1) avec rupture à l'instant $k^* = 500$; le paramètre $\phi_1 = 0.4$ change en 0.2. c-), d-), e-) and f-) les statistiques $\widehat{Q}^{(1)}_{n,k}$ and $\widehat{Q}^{(2)}_{n,k}$ correspondantes. La droite horizontale délimite la zone critique.

La Figure 1.5 c-) et d-) montrent que lorsqu'il n'y a pas rupture, les statistiques $\widehat{Q}^{(1)}_{n,k}$ et

$\widehat{Q}_{n,k}^{(2)}$ sont très faibles par rapport à la valeur critique. La Figure 1.5 e-) et f-) montrent que ces statistiques prennent des grandes valeurs sous $\mathbf{H}_1$ et atteignent leurs maximum au voisinage de l'instant où la rupture s'est produite. Le Théorème 1.8 montre que le maximum de ces deux statistiques diverge vers l'infini.

Théorème 1.8. *Sous $\boldsymbol{H}_1$, supposons que les hypothèses $\boldsymbol{D}(\Theta)$, **Id**(Θ), **Var**, $\boldsymbol{C}$ et $\boldsymbol{K}(f_\theta, M_\theta, \Theta)$ sont satisfaites. Si, $\theta_1^*, \theta_K^* \in \overset{\circ}{\Theta}(4)$ et $\theta_1^* \neq \theta_K^*$, alors*

$$\widehat{Q}_n \xrightarrow[n\to\infty]{a.s.} \infty.$$

Le Théorème 1.8 montre que la procédure est consistante en puissance.

1.4.4 Procédure d'estimation de rupture multiple

La procédure décrite ci-dessus permet de tester la présence ou non des ruptures dans les observations. Pour estimer les instants de ruptures, on va joindre à cette procédure, un algorithme de type ICSS (Itereted Cumulative Sums of Squares) développé par Inclán et Tiao [37] en 1994.

Algorithme ICSS pour l'estimation de rupture multiple

- **Etape 0** Poser $t_0 = 0$.
- **Etape 1** Effectuer le test ci-dessus sur les observations $X_{t_0}, \cdots, X_n$. Notons $k(t_0, n)$ le point où la statistique $\widehat{Q}_n$ atteint sa valeur maximale.
  - Si H_0 est acceptée, alors il n'y a pas de rupture apparente dans les observations. L'algorithme s'arrête ;
  - sinon, passer à l'Etape 2a.
- **Etape 2a** Poser $t_1 = k(t_0, n)$. Effectuer le test sur les observations $X_{t_0}, \cdots, X_{t_1}$. Si H_0 est rejetée, il y a un nouveau point de rupture potentiel ; dans ce cas, répéter l'Etape 2a jusqu'à ce que H_0 soit acceptée. Ainsi, on conclut qu'il n'y a pas de rupture apparente dans les observations $X_{t_0}, \cdots, X_{t_1}$ et le premier instant de rupture est $t_{first} = t_1$.
- **Etape 2b** Poser $t_0 = k(t_0, n)$. Procéder comme à l'Etape 2a pour chercher le dernier instant de rupture en partant des observations $X_{t_0}, \cdots, X_n$. Supposons que ce dernier instant de rupture trouvé est $t_{last} = t_0 - 1$.
- **Etape 2c** Si $t_{first} = t_{last}$, alors il y a une seule rupture et l'algorithme s'arrête. Sinon, prendre t_{first} et t_{last} comme des instants de rupture ; poser $t_0 = t_{first}$ et $n = t_{last}$ et répéter l'Etape 1 et l'Etape 2 en partant des observations $X_{t_0}, \cdots, X_n$.
- **Etape 3** Notons $t_1, \cdots, t_K$ $(K \geq 1)$ les instants de ruptures obtenus ci-dessus. On passe à la validation *i.e.* poser $t_0 = 0$ et $t_{K+1} = n$; pour $j = 1, \cdots, K$, effectuer le test ci-dessus sur les observations $X_{t_{j-1}+1}, \cdots, X_{t_{j+1}}$. Si H_0 est rejetée, garder le point t_j comme instant de rupture sinon, le supprimer. Répéter l'Etape 3 jusqu'à ce que le nombre de rupture soit constant entre deux itérations.

Cet algorithme est utilisé au Chapitre 5 pour la détection des ruptures dans les débits du bassin versant de la Sanaga.

1.5 Détection on-line de rupture

Dans les deux cas traités ci-dessus, on détecte les ruptures dans la trajectoire du processus qu'on a observé. Mais, dans beaucoup de systèmes industriels, les données arrivent au fur et à mesure que le système fonctionne et l'on souhaite détecter les ruptures aussitôt qu'elles se produisent. C'est le cas par exemple du contrôle de qualité en entreprise. Ce problème est la détection on-line des ruptures que nous allons maintenant traiter.

1.5.1 Présentation du problème

On considère un processus $X = (X_t)_{t\in\mathbb{Z}}$ appartenant à la famille de modèle $\mathcal{M}_T(f_\theta, M_\theta)$ (définit en (1.2)) avec $\theta \in \Theta$. On suppose que les observations $(X_1, \dots, X_n)$ sont disponibles et dépendent d'un seul paramètre *i.e.* il existe $\theta_0^* \in \Theta$ tel que $(X_1, \cdots, X_n)$ appartient à la classe $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0^*}, f_{\theta_0^*})$. Dans la suite, des nouvelles valeurs $X_{n+1}, X_{n+2} \cdots, X_k, X_{k+1}, \cdots$ seront observées. Pour chaque nouvelle observation, on veut savoir si c'est le modèle dépendant toujours du paramètre θ_0^* qui l'a générée. Plus précisément, on considère le test suivant :
H_0 : θ_0^* est constant dans la suite des observations $X_1, \cdots, X_n, X_{n+1}, \cdots$ *i.e.* les observations $X_1, \cdots, X_n, X_{n+1}, \cdots$ appartiennent à la classe $\mathcal{M}_{\mathbb{N}}(M_{\theta_0^*}, f_{\theta_0^*})$;

H_1 : il existe $k^* > n$, $\theta_1^* \in \Theta$ tel que $X_1, \cdots, X_n, X_{n+1}, \cdots, X_{k^*}$ appartiennent à la classe $\mathcal{M}_{\{1,\cdots,k^*\}}(M_{\theta_0^*}, f_{\theta_0^*})$ et $X_{k^*+1}, X_{k^*+2}, \cdots$ appartiennent à $\mathcal{M}_{\{k^*+1,\cdots,\}}(M_{\theta_1^*}, f_{\theta_1^*})$.

L'objectif est de développer une procédure permettant de tester ces deux hypothèses. On verra dans la suite que l'enjeu c'est aussi de réduire autant que possible le délai de détection.

1.5.2 Nouvelle procédure de détection on-line de rupture

Dans la suite, on suppose que les observations $(X_1, \dots, X_n)$ appartenant à la classe $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0^*}, f_{\theta_0^*})$ sont disponibles. La nouvelle procédure est décrite par une statistique $\widehat{C}_{k,\ell}$ (appelée détecteur) basée sur les observations $X_1, \cdots, X_k$ avec $k > n$. Pour $1 \leq \ell \leq \ell'$, posons

$$T_{\ell,\ell'} := \{\ell, \ell+1, \cdots, \ell'\}.$$

D'après [40], la matrice $\widehat{G}(T_{1,n})$ (voir 1.12) est symétrique et asymptotiquement définie positive. Ainsi, $\widehat{G}(T_{1,n})^{-1/2}$ existe pour n "grand".

Dans la suite, $\widehat{F}(T_{1,n})$ désigne la matrice définie en 1.13 et $(v_n)_{n\in\mathbb{N}}$ désigne une suite des entiers positifs satisfaisants

$$v_n \leq n/2, \quad v_n \to \infty \quad \text{et} \quad v_n/\sqrt{n} \to 0 \ \ (n \to \infty).$$

Posons

$$\Pi_{n,k} := \{n - v_n, n - v_n + 1, \cdots, k - v_n\}.$$

et définissons le détecteur pour tout $k > n$ et $\ell \in \Pi_{n,k}$

$$\widehat{C}_{k,\ell} := \sqrt{n}\,\frac{k-\ell}{k}\,\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})\big)\big\|.$$

Notre procédure consiste à rejeter H_0 dès le premier instant $k > n$ tel qu'il existe $\ell \in \Pi_{n,k}$ satisfaisant $\widehat{C}_{k,\ell} > c$ pour une constante $c > 0$ fixée.

Pour la généralité, nous allons définir la valeur critique à partir d'une fonction $b : (0,\infty) \mapsto (0,\infty)$, que nous supposons (dans toute la suite) décroissante, continue et satisfaisant $\operatorname*{Inf}_{0<t<\infty} b(t) > 0$. Ainsi, la procédure on-line s'arrête (et on déclare l'existence d'une rupture) dès le premier instant $k > n$ tel qu'il existe $\ell \in \Pi_{n,k}$ satisfaisant $\widehat{C}_{k,\ell} > b((k-\ell)/n)$. Définissons le temps d'arrêt

$$\tau(n) := \operatorname{Inf}\{k > n \,/\, \exists \ell \in \Pi_{n,k},\ \widehat{C}_{k,\ell} > b((k-\ell)/n)\} = \operatorname{Inf}\{k > n \,/\, \max_{\ell \in \Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\}.$$

Il s'ensuit que

$$\begin{aligned} P\{\tau(n) < \infty\} &= P\Big\{\text{qu'il existe } k > n \text{ tel que } \max_{\ell \in \Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\} \\ &= P\Big\{\sup_{k>n} \max_{\ell \in \Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\}. \end{aligned} \tag{1.14}$$

L'objectif est donc de trouver une fonction convenable b telle que pour un seuil donné $\alpha \in]0,1[$ on a

$$\lim_{n\to\infty} P_{H_0}\{\tau(n) < \infty\} = \alpha$$

et

$$\lim_{n\to\infty} P_{H_1}\{\tau(n) < \infty\} = 1.$$

1.5.3 Résultats asymptotiques

Sous l'hypothèse $\mathbf{H}_0$

Le Théorème 1.9 donne la loi limite sous H_0.

Théorème 1.9. *On suppose que les hypothèses* $\boldsymbol{D}(\Theta)$, $\boldsymbol{Id}(\Theta)$, $\boldsymbol{Var}(\Theta)$, $\boldsymbol{K}(f_\theta, M_\theta, \Theta)$ *sont satisfaites. Sous l'hypothèse* H_0, *si* $\theta_0^* \in \overset{\circ}{\Theta}(4)$ *alors*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1} \sup_{1<s<t} \frac{\|W_d(s) - sW_d(1))\|}{t\ b(s)} > 1\Big\}$$

où W_d *est un pont brownien de dimension* d.

Le Corollaire 1.10 permet de construire la valeur critique du test lorsque b est constante.

Corollaire 1.10. *On suppose* $b(t) = c > 0$ *pour tout* $t \geq 0$. *Sous les hypothèses du Théorème 1.9,*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1} \sup_{1<s<t} \frac{1}{t} \|W_d(s) - sW_d(1))\| > c\Big\} = P\{U_d > c\}$$

où $U_d = \sup_{0<u<1} f(u)\,\|W_d(u)\|$ *avec* $f(u) = \dfrac{\sqrt{9-u}+\sqrt{1-u}}{\sqrt{9-u}+3\sqrt{1-u}} \Big(\dfrac{2}{3-u+\sqrt{(9-u)(1-u)}}\Big)^{1/2}$.

Ainsi, à un seuil $\alpha \in]0,1[$, prendre $c = c(\alpha)$ le quantile d'ordre $(1-\alpha)$ de la loi de U_d. Ces quantiles peuvent se calculer par la méthode de Monte-Carlo. Nous avons fait les calculs pour $d \in \{1, \cdots, 5\}$ (voir Chapitre 4).

Sous l'hypothèse $\mathbf{H}_1$

On suppose que le paramètre θ_0^* change en θ_1^* à l'instant $k^* > n$, avec $\theta_1^* \in \Theta$ et $\theta_0^* \neq \theta_1^*$. Alors

Théorème 1.11. *On suppose que les hypothèses* $\boldsymbol{D}(\Theta)$, ***Id***(Θ), ***Var***(Θ), $\boldsymbol{K}(f_\theta, M_\theta, \Theta)$ *sont satisfaites et que* $\sup_{0<t<\infty} b(t) < \infty$. *Sous l'hypothèse* H_1, *si* $\theta_0^*, \theta_1^* \in \overset{\circ}{\Theta}(4)$ *avec* $\theta_1^* \neq \theta_0^*$ *alors pour* $k^* = k^*(n) > n$ *tel que* $\limsup_{n\to\infty} k^*(n)/n < \infty$ *et* $k_n = k^*(n) + n^\delta$ *avec* $\delta \in (1/2, 1)$,

$$\max_{\ell \in \Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} \xrightarrow[n\to\infty]{a.s.} \infty.$$

Le Corollaire 1.12 est immédiat à partir de la relation (1.14).

Corollaire 1.12. *Sous les hypothèses du Théorème 1.11,*

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = 1.$$

1.5.4 Illustration de la procédure

Si $b \equiv c > 0$ est une fonction constante, alors

$$P\{\tau(n) < \infty\} = P\Big\{\sup_{k>n} \max_{\ell \in \Pi_{n,k}} \widehat{C}_{k,\ell} > c\Big\}.$$

Posons alors

$$\widehat{C}_k = \max_{\ell \in \Pi_{n,k}} \widehat{C}_{k,\ell} \quad \text{pour tout} \quad k > n.$$

On considère un processus GARCH(1,1) : $X_t = \sigma_t \xi_t$ avec $\sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$. Les données historiques (observations disponibles) $X_1, \cdots, X_{500}$ dépendent du paramètre $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*)$. Sur la Figure 1.6 a-), le paramètre $\theta_0^* = (0.01, 0.3, 0.2)$ reste constant dans la suite des observations $X_{501}, \cdots, X_{1000}$. Sur la Figure 1.6 b-), le paramètre $\theta_0^* = (0.01, 0.3, 0.2)$ change en $\theta_1^* = (0.05, 0.5, 0.2)$ à l'instant $k^* = 750$.
La 1.6 a-) montre que sous H_0, la statistique $\widehat{C}_k$ est très inférieure à la valeur critique du test. D'après, le Corollaire 1.12, cette statistique restera (asymptotiquement) en dessous de la valeur critique dans 95% des cas. La Figure 1.6 b-) montre une forte croissance de la statistique $\widehat{C}_k$ sous H_1 dès que l'instant de rupture est dépassé. Le Théorème 1.11 garantit que la valeur critique sera presque sûrement (et asymptotiquement quand $n \to \infty$) dépassée.

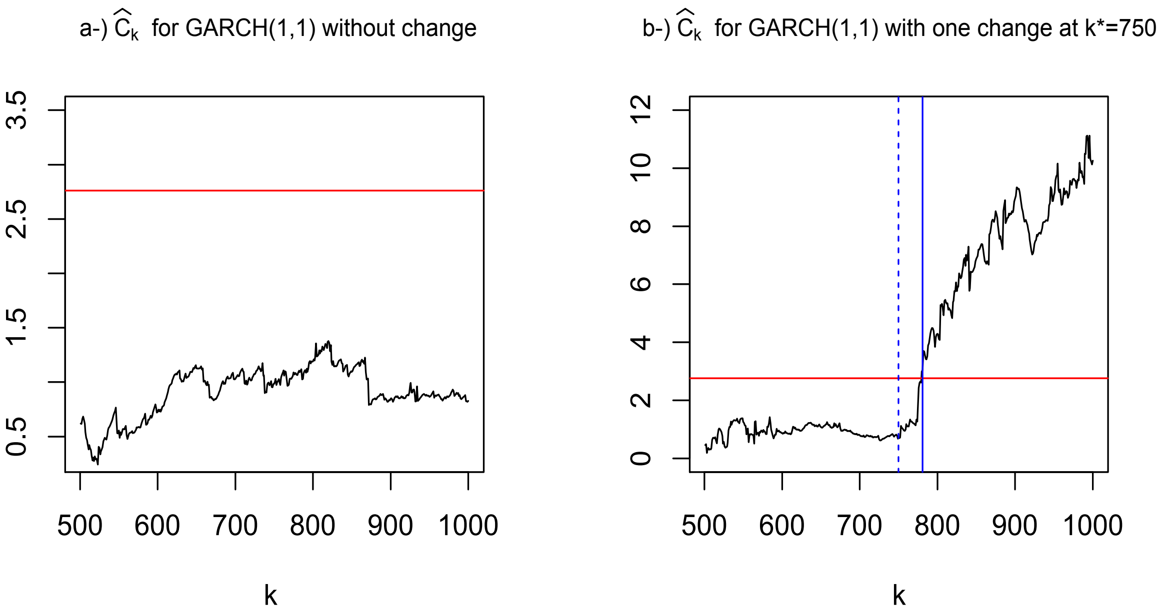

Figure 1.6 – Réalisation de la statistique $\widehat{C}_k$ pour un processus GARCH(1,1). La droite horizontale représente la valeur critique du test, la droite verticale en pointillée représente le vrai instant de rupture et la droite verticale représente l'instant où la rupture est détectée.

Chapitre 2

Multiple breaks detection in general causal time series using penalized quasi-likelihood

Abstract

This paper is devoted to the off-line multiple breaks detection for a general class of models. The observations are supposed to fit a parametric causal process (such as classical models AR(∞), ARCH(∞) or TARCH(∞)) with distinct parameters on multiple periods. The number and dates of breaks, and the different parameters on each period are estimated using a quasi-likelihood contrast penalized by the number of distinct periods. For a convenient choice of the regularization parameter in the penalty term, the consistency of the estimator is proved when the moment order r of the process satisfies $r \geq 2$. If $r \geq 4$, the length of each approximative segment tends to infinity at the same rate as the length of the true segment and the parameters estimators on each segment are asymptotically normal. Compared to the existing literature, we added the fact that a dependence is possible over distinct periods. To be robust to this dependence, the chosen regularization parameter in the penalty term is larger than the ones from BIC approach. We detail our results which notably improve the existing ones for the AR(∞), ARCH(∞) and TARCH(∞) models. The procedure is implemented using the slope estimation of the regularization parameter and a dynamic programming algorithm. It is an $O(n^2)$ complexity algorithm that we apply on AR(1), GARCH$(1,1)$ and TARCH(1) processes and on the FTSE index data.

Keywords : Change detection, Causal processes, ARCH(∞) processes, AR(∞) processes, Quasi-maximum likelihood estimator, Model selection by penalized likelihood.

Note

The content of this chapter is based on a paper, written in collaboration with Jean-Marc Bardet and Olivier Wintenberger, published in the Electronic Journal of Statistics.

2.1 Introduction

The breaks detection is a classical problem as well as in the statistic than in the signal processing community. The first important result in this topic was obtained by Page [61] in 1955 and real advances have been done during the seventies, notably with the results of Hinkley (see for instance [31]) and the break detection became a distinct and important area of research in statistic (see the book of Basseville and Nikiforov [11] for a large overview).
Two approaches are generally considered for solving a problem of breaks detection : an 'on-line' approach leading to sequential estimation and an 'off-line' approach when the series of observations is complete. Concerning this last approach, numerous results were obtained for independent random variables in a parametric frame (see for instance Bai and Perron [7]). The case of the off-line detection of multiple change-points in a parametric or semiparametric frame for dependent variables or time series also provided an important literature. The present paper is a new contribution to this problem.

In this paper, we consider the following change-point problem : for $j = 1, 2, \cdots, K^*$,

$$X_t = g_{\theta_j^*}(\xi_t, X_{t-1}, X_{t-2}, \cdots) \quad \text{for all } t \in \{t_{j-1}^* + 1, t_{j-1}^* + 2, \cdots, t_j^*\} \tag{2.1}$$

where g_θ is a parametric function satisfying assumptions detailed in Section 2.2, $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) $\mathbb{R}^p$-random vectors called the innovations, $K^* - 1 \in \mathbb{N}$ is the unknown number of the breaks, $t_0^* = 0 < t_1^* < \ldots < t_{K^*-1}^* < n = t_{K^*}^*$ with $(t_j^*)_{1\leq j\leq K^*-1} \in \mathbb{N}$ are the $K^* - 1$ unknown dates of the breaks, $\theta_j^* \in \Theta \subset \mathbb{R}^d$ for $j = 1, \ldots, K^*$ are the unknown parameters of the model. Note that the assumptions on g_θ are weaker enough for X to be for instance AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes on each period.
The aim of our statistical procedure is the estimation of the unknown parameters $\big(K^*, (t_j^*)_{1\leq j\leq K^*-1}, (\theta_j^*)_{1\leq j\leq K^*}\big)$ in the problem (2.1). In the literature, it is generally supposed that X is a stationary process on each set $\{t_{j-1}^* + 1 \cdots, t_j^*\}$ and is independent on each $\{t_{i-1}^* + 1, \cdots, t_i^*\}$ of the other $\{t_{k-1}^* + 1, \cdots, t_k^*\}$, $k \neq i$ (for instance in [51], [45] and [25]). Here the problem (2.1) does not induce such assumptions and thus the framework is closer to the applications, see Remark 1 in [25].

In the problem of change-points detection, numerous papers were devoted to the CUSUM procedure (see for instance Kokozska and Leipus [45] in the specific case of ARCH(∞) processes). In Lavielle and Ludena [50] a "Whittle" contrast is used for estimating the breaks dates in the spectral density of piecewise long-memory processes (in a semi-parametric framework). Davis *et al.* [24] proposed a likelihood ratio as the estimator of breaks for an AR(p) process. Lavielle and Moulines [51] consider a general contrast using the mean square errors for estimating the parameters. In Davis *et al.* [25], the criteria called Minimum Description Length (MDL) is applied to a large class of nonlinear time series.
We consider here a semiparametric estimator based on a penalized contrast (so-called *penQLIK* in the sequel) using the quasi-likelihood function. For usual stationary time series, the conditional quasi-likelihood (so-called $QLIK$ in the sequel) is constructed as follow :

1. Assume the process $(\xi_t)_{t\in\mathbb{Z}}$ is a Gaussian sequence and compute the conditional likelihood (with respect to $\sigma\{X_0, X_{-1}, \ldots\}$) based on the unobservable infinite realization of $(X_t)_{t\in\mathbb{Z}}$;
2. Approximate this computation for a sample $(X_1, \ldots, X_n)$;

3. Apply this approximation even if the process of the innovations is not a Gaussian sequence.

The quasi-maximum likelihood estimator (QMLE) obtained by maximizing the $QLIK$ has convincing asymptotic properties in the case of GARCH processes (see Jeantheau [38], Berkes *et al.* [13], Franck and Zakoian [29]) or generalizations of GARCH processes (see Mikosch and Straumann [68], Robinson and Zaffaroni [65]). Bardet and Wintenberger [8] study the asymptotic normality of the QMLE of θ applied to the class of models considered here. Thus, when K^* is known, a natural estimator of the parameter $(\underline{t}^*, \underline{\theta}^*) = \big((t_j^*)_{1\leq j\leq K^*-1}, (\theta_j^*)_{1\leq j\leq K^*}\big)$ for a process satisfying (2.1) is the QMLE on every intervals $[t_j+1,\ldots,t_{j+1}]$ and every parameters θ_j for $1 \leq j \leq K^*$. However we consider here that K^* is unknown and such method cannot be directly used. The chosen solution is to penalize the contrast by an additional term $\kappa_n K$, where the regularization parameters κ_n form an increasing sequence of real numbers (see the final expression of the penalized contrast in (2.4)). Such procedure of penalization was previously used for instance by Yao [69] to estimate the number of change-points with the Schwarz criterion and by Lavielle and Moulines [51]. Hence the minimization of the penalized contrast leads to an estimator (see (2.5)) of the parameters $(K^*, \underline{t}^*, \underline{\theta}^*)$.
Classical heuristics such as the BIC one lead to choose $\kappa_n \propto \log n$. In our study, such penalties terms are excluded in some cases, when the models in (2.1) are very dependent on their whole past, see Section 2.3 (and simulation results) for more details. Roughly speaking, an explanation of this can be provided by the simple relation :

$$\begin{aligned} penQLIK(K,\underline{t},\underline{\theta}) &= QLIK(K,\bar{t},\underline{\theta}) + \kappa_n K \\ &= \Big(QLIK(K,\underline{t},\underline{\theta}) - \widetilde{QLIK}(K,\underline{t},\underline{\theta})\Big) + \widetilde{QLIK}(K,\underline{t},\underline{\theta}) + \kappa_n K \end{aligned}$$

where $\widetilde{QLIK}$ is the conditional quasi-likelihood of a process following (2.1) except that it is composed by stationary time series on each period which are independent of the stationary processes defined on the other periods. Using moment bounds we will prove in Section 2.6 that $\big|QLIK(K,\underline{t},\underline{\theta}) - \widetilde{QLIK}(K,\underline{t},\underline{\theta})\big| = O_P(u_n)$ with $u_n \to \infty$ and $u_n/n \to 0$, where $(u_n)_{n\in\mathbb{N}}$ depends on the Lipshitzian behavior of g_θ. Since $\widetilde{QLIK}(K,\underline{t},\underline{\theta}) \sim C\,n \;\; a.s.$ when $n \to \infty$ from results obtained in [8], it is clear that the penalty term can play a role only if $\kappa_n >> u_n$. Finally, we will show that under weak conditions on the model, the regularization parameter $\kappa_n \propto \sqrt{n}$ over-penalizes the number of breaks for avoiding artificial breaks in cases of models very dependent on their whole past (see Section 2.3 for details). Such a choice of κ_n is robust to the (possibly strong) dependence.

The main results of the paper are the following : the estimator $\big(\widehat{K}_n, (\widehat{t}_j/n)_{1\leq j\leq \widehat{K}_n-1}, (\widehat{\theta}_j)_{1\leq j\leq \widehat{K}_n}\big)$ is consistent under Lipshitzian condition on g_θ and when the moments of order $r \geq 2$ of the innovations and X are finite. If moreover Lipshitzian conditions are also satisfied by the derivatives of g_θ and if $r \geq 4$, then the convergence rate of $(\widehat{t}_j/n)_{1\leq j\leq \widehat{K}_n-1}$ is $O_P(w_n)$ for any sequence $(w_n)_n$ such that $w_n >> n^{-1}$ and a Central Limit Theorem (CLT) for $(\widehat{\theta}_j)_{1\leq j\leq \widehat{K}_n}$ with a $\sqrt{n}$-convergence rate is established. These results are "optimal" in the sense that the convergence rate is the same than in an independent setting.

After detailing the particular cases of AR(∞), ARCH(∞) and TARCH(∞) satisfying the break-point problem (2.1), the estimator is applied to generated trajectories of such time series. Two difficulties appeared. Firstly, the computation time was very long and exponentially increased with K. We solved this problem by using a dynamic programming

algorithm which is a $O(n^2)$ complexity algorithm. We also considered only small length trajectories ($n \leq 2000$). Secondly, we obtained the consistency of the estimator of K^* as a theoretical result in all considered model regardless of their dependence properties when $\kappa_n \propto \sqrt{n}$ when $n \to \infty$. We will see that for particular models such that ARMA(p,q) or GARCH(p,q) a BIC-type penalty with $\kappa_n \propto \log n$ is also possible, but $\kappa_n \propto \sqrt{n}$ ensures the convergence for a larger class of models (including AR(∞), ARCH(∞) or TARCH(∞) processes).
However, for n not too large (for instance $n = 1000$) the choice of $\kappa_n = \sqrt{n}$ very often led to $\widehat{K}_n \neq K^*$. Hence we chose to implement a data-driven procedure for estimating κ_n (denoted $\hat{\kappa}_n$ in the sequel) using a slope estimation method (see [12]), such procedure being nowadays often used in the model selection frame. In such a way, the results of simulations are clearly satisfying (see Section 4.4). The estimation procedure is also applied to financial data and this provides estimating dates of breaks corresponding with key dates of financial crisis.

The following Section 2.2 is devoted to the assumptions and the study of the existence of a nonstationary solution of the change point problem (2.1). The definition of the estimator and its asymptotic properties are studied in Section 2.3. The particular examples of AR(∞), ARCH(∞) and TARCH(∞) processes are detailed in Section 2.4, while the concrete estimation procedure and numerical applications are presented in Section 4.4. Finally, Section 2.6 contains the main proofs.

2.2 Assumptions and existence of a non-stationary solution

2.2.1 Notation and assumptions

Let $\theta \in \mathbb{R}^d$ and M_θ and f_θ be real-valued measurable functions such that for all $(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $M_\theta\big((x_i)_{i\in\mathbb{N}}\big) \neq 0$. In this paper, we consider a general class $\mathcal{M}_T(M_\theta, f_\theta)$ of causal (non-anticipative) time series. Let $T \subset \mathbb{Z}$ and $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) random variables called the innovations and satisfying $\mathrm{var}(\xi_0) = 1$. Define

Class $\mathcal{M}_T(M_\theta, f_\theta)$: *The process $X = (X_t)_{t\in\mathbb{Z}}$ belongs to $\mathcal{M}_T(M_\theta, f_\theta)$ if it satisfies the relation :*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{2.2}$$

The existence and properties of these general affine processes were studied in Bardet and Wintenberger [8] as a particular case of chains with infinite memory considered in Doukhan and Wintenberger [27]. Numerous classical real valued time series are included in $\mathcal{M}_{\mathbb{Z}}(M, f)$: for instance AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes.

For obtaining conditions of existence of a process included in $\mathcal{M}_T(M_\theta, f_\theta)$ first define the following different norms :

1. $\|\cdot\|$ applied to a vector denotes the Euclidean norm of the vector ;
2. for any compact set $\mathcal{K} \subseteq \mathbb{R}^d$ and for any $g : \mathcal{K} \longrightarrow \mathbb{R}^{d'}$; $\|g\|_{\mathcal{K}} = \sup_{\theta\in\mathcal{K}}(\|g(\theta)\|)$;
3. for all $x = (x_1, \cdots, x_K) \in \mathbb{R}^K$, $\|x\|_m = \max\limits_{i=1,\cdots,K} |x_i|$;
4. if Y is a random vvector with finite r-order moments, we set $\|Y\|_r = (\mathbb{E}\|Y\|^r)^{1/r}$.

Let $\Psi_\theta = M_\theta,\ f_\theta$ and $i = 0,\ 1,\ 2$, then for any compact set $\mathcal{K} \subseteq \mathbb{R}^d$, define

Assumption $\mathbf{A}_i(\Psi_\theta, \mathcal{K})$: *Assume that* $\|\partial^i \Psi_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *and there exists a sequence of non-negative real number* $(\alpha_k^{(i)}(\Psi_\theta, \mathcal{K}))_{k\geq 1}$ *such that* $\sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K}) < \infty$ *satisfying*

$$\Big\|\frac{\partial^i \Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i \Psi_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K})|x_k - y_k| \quad \textit{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

In the sequel we refer to the particular case called "ARCH-type process", if $f_\theta = 0$ and the following assumption holds on $h_\theta = M_\theta^2$:

Assumption $\mathbf{A}_i(h_\theta, \mathcal{K})$: *Assume that* $f_\theta = 0$, $\|\partial^i h_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *and there exists a sequence of non-negative real number* $(\alpha_k^{(i)}(h_\theta, \mathcal{K}))_{k\geq 1}$ *such that* $\sum_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \mathcal{K}) < \infty$ *satisfying*

$$\Big\|\frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \mathcal{K})|x_k^2 - y_k^2| \quad \textit{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

Now, for any $i = 0, 1, 2$ and any compact $\mathcal{K} \subset \mathbb{R}^d$, under Assumptions $A_i(f_\theta, \mathcal{K})$ and $A_i(M_\theta, \mathcal{K})$, denote :

$$\beta^{(i)}(\mathcal{K}) := \sum_{k\geq 1} \beta_k^{(i)}(\mathcal{K}) \quad \text{where} \quad \beta_k^{(i)}(\mathcal{K}) := \alpha_k^{(i)}(f_\theta, \mathcal{K}) + (\mathbb{E}|\xi_0|^r)^{1/r} \alpha_k^{(i)}(M_\theta, \mathcal{K}),$$

and under Assumption $A_i(h_\theta, \mathcal{K})$

$$\tilde{\beta}^{(i)}(\mathcal{K}) := \sum_{k\geq 1} \tilde{\beta}_k^{(i)}(\mathcal{K}) \quad \text{where} \quad \tilde{\beta}_k^{(i)}(\mathcal{K}) := (\mathbb{E}|\xi_0|^r)^{2/r} \alpha_k^{(i)}(h_\theta, \mathcal{K}).$$

The dependence with respect to r of the coefficients $\beta^{(i)}$ and $\tilde{\beta}^{(i)}$ are omitted for notational convenience. From now on let us fix Θ a compact subset of $\mathbb{R}^d$ satisfying some contraction properties :

Assumption A : *Assume there exists* $r \geq 1$ *such that for all* $\theta \in \Theta$ *either* $A_0(f_\theta, \{\theta\})$ *and* $A_0(M_\theta, \{\theta\})$ *hold with* $\beta^{(0)}(\{\theta\}) < 1$ *either* $f_\theta = 0$ *and* $A_0(h_\theta, \{\theta\})$ *holds with* $\tilde{\beta}^{(0)}(\{\theta\}) < 1$.

From [8] we have :

Proposition 2.2.1. If $\theta \in \Theta$ satisfies A for some $r \geq 1$, there exists a unique causal (non anticipative, i.e. X_t is independent of $(\xi_i)_{i>t}$ for $t \in \mathbb{Z}$) solution $X = (X_t)_{t\in\mathbb{Z}} \in \mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$ which is stationary, ergodic and satisfies $\|X_0\|_r < \infty$.

The assumption A is classical when studying the existence of stationary solution of general models. For instance, Duflo [28] used such a Lipschitz-type inequality to show the existence of Markov chains. The elements of the compact set Θ satisfies one Lipschitz-type condition specified either for general causal models either for ARCH-type models. This distinction is adequate as for ARCH-type models $A_0(h_\theta, \{\theta\})$ is less restrictive than $A_0(M_\theta, \{\theta\})$. Remark that assumption $\tilde{\beta}^{(0)}(\theta) < 1$ is optimal for the stationarity of order $r \geq 1$ but not for the strict stationarity of the solution of an ARCH-type model.

Let $\theta \in \Theta$ and $X = (X_t)_{t\in\mathbb{Z}}$ a stationary solution included in $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$. For studying QMLE properties, it is convenient to assume the following assumptions :

Assumption D(Θ) **:** $\exists \underline{h} > 0$ such that $\inf\limits_{\theta\in\Theta}(|h_\theta(x)|) \geq \underline{h}$ for all $x \in \mathbb{R}^{\mathbb{N}}$.

Assumption Id(Θ) **:** For all $\theta, \theta' \in \Theta^2$,

$$\Big(f_\theta(X_0, X_{-1}, \cdots) = f_{\theta'}(X_0, X_{-1}, \cdots) \text{ and } h_\theta(X_0, X_{-1}, \cdots) = h_{\theta'}(X_0, X_{-1}, \cdots) \text{ a.s.}\Big) \Rightarrow \theta = \theta'.$$

Assumption Var(Θ) **:** For all $\theta \in \Theta$, one of the families $\big(\dfrac{\partial f_\theta}{\partial \theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ or $\big(\dfrac{\partial h_\theta}{\partial \theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ is a.e. linearly independent.

Assumption D(Θ) will be required to define the QMLE, Id(Θ) to show the consistence of the QMLE and Var(Θ) to show the asymptotic normality.

2.2.2 Existence of the solution to the change-point problem

Using the class $\mathcal{M}_T(M_\theta, f_\theta)$, the problem (2.1) of change-point detection can be formulated as follows : assume that a trajectory $(X_1, \cdots, X_n)$ of $X = (X_t)_{t\in\mathbb{Z}}$ is observed where

$$X \in \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*}) \quad \text{for all } j = 1, \ldots, K^*, \qquad \text{with} \tag{2.3}$$

- $K^* \in \mathbb{N}^*$, $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \ldots, t_j^*\}$ with $0 < t_1^* < \ldots < t_{K^*-1}^* < n$, $t_j^* \in \mathbb{N}$ and by convention $t_0^* = -\infty$ and $t_{K^*}^* = \infty$;
- $\theta_j^* = (\theta_{j,1}^*, \cdots, \theta_{j,d}^*) \in \Theta \subset \mathbb{R}^d$ for $j = 1, \ldots, K^*$.

Consider the problem (2.3). Then the past of X before the time $t = 0$ depends on θ_1^* and the future after $t = n$ depends on $\theta_{K^*}^*$. The number $K^* - 1$ of breaks, the instants $t_1^*, \cdots, t_{k^*-1}^*$ of breaks and parameters $\theta_1^*, \cdots, \theta_{K^*}^*$ are unknown. Consider first the following notation.

Notation .

- *For* $K \geq 2$, $\mathcal{F}_K = \{\underline{t} = (t_1, \ldots t_{K-1})\ ;\ 0 < t_1 < \ldots < t_{K-1} < n\}$. *In particular,* $\underline{t}^* = (t_1^*, \ldots, t_{K^*-1}^*) \in \mathcal{F}_{K^*}$ *is the true vector of instants of change ;*
- *For* $K \in \mathbb{N}^*$ *and* $\underline{t} \in \mathcal{F}_K$, $T_k = \{t \in \mathbb{Z},\ t_{k-1} < t \leq t_k\}$ *and* $n_k = Card(T_k)$ *with* $1 \leq k \leq K$. *In particular ;* $T_j^* = \{t \in \mathbb{Z},\ t_{j-1}^* < t \leq t_j^*\}$ *and* $n_j^* = Card(T_j^*)$ *for* $1 \leq j \leq K^*$. *For all* $1 \leq k \leq K$ *and* $1 \leq j \leq K^*$, *let* $n_{kj} = Card(T_j^* \cap T_k)$ *;*

The following proposition establishes the existence of the non stationary solution of the problem (2.3) and its moments properties.

Proposition 2.2.2. Consider the problem (2.3) with $\theta_j^* \in \Theta$ for all $j = 1, \ldots, K^*$, Θ satisfying A for some $r \geq 1$. Then

(i) there exists a solution $X = (X_t)_{t\in\mathbb{Z}}$ of the model (2.3) and X is a causal time series.

(ii) there exists a constant $C > 0$ such that for all $t \in \mathbb{Z}$ we have $\|X_t\|_r \leq C$.

The problem (2.3) distinguishes the case $t \in T_1^* = \{1, \ldots, t_1^*\}$ to the other ones since it is easy to see that $(X_t)_{t\in T_1^*}$ is a stationary process while $(X_t)_{t>t_1^*}$ is not. However, all the results of this paper hold if $(X_t)_{t\in T_1^*}$ is defined as the other $(X_t)_{t\in T_j^*}$, $j \geq 2$ (by defining a break in $t = 0$ setting $X_t = 0$ for $t \leq 0$ for instance).

2.3 The estimation procedure and the asymptotic behavior of the estimator

2.3.1 The penalized QLIK contrast

The estimation procedure of the number of breaks $K^* - 1$, the instants of breaks $\underline{t}^*$ and the parameters $\underline{\theta}^*$ is based on the minimum of a penalized QLIK contrast. By definition, if $X \in \mathcal{M}_T(f_\theta, M_\theta)$ then the conditional (to the past values of X) mean and the variance are given by, respectively, $f_\theta(X_{s-1}, \ldots)$ and $h_\theta(X_{s-1}, \ldots)$. Therefore, with the notation $f_\theta^s = f_\theta(X_{s-1}, X_{s-2} \ldots)$, $M_\theta^s = M_\theta(X_{s-1}, X_{s-2} \ldots)$ and $h_\theta^s = M_\theta^{s2}$, we deduce the quasi-likelihood of X on a period T :

$$L_n(T, \theta) := -\frac{1}{2} \sum_{s \in T} q_s(\theta) \;\text{ with }\; q_s(\theta) := \frac{(X_s - f_\theta^s)^2}{h_\theta^s} + \log(h_\theta^s) \, .$$

By convention, we set $L_n(\emptyset, \theta_k) := 0$. Since only $X_1, \ldots, X_n$ are observed, $L_n(T, \theta)$ cannot be computed because it depends on the past values $(X_{-j})_{j \in \mathbb{N}}$. We approximate it by the QLIK criteria on a period T :

$$\widehat{L}_n(T, \theta) := -\frac{1}{2} \sum_{s \in T} \widehat{q}_s(\theta) \quad \text{where} \quad \widehat{q}_s(\theta) := \frac{(X_s - \widehat{f}_\theta^s)^2}{\widehat{h}_\theta^s} + \log(\widehat{h}_\theta^s)$$

with $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \ldots, X_1, u)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \ldots, X_1, u)$ and $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$ for any deterministic sequence $u = (u_n)$ with finitely many non-zero values.

Remark 2.3.1. For convenience, in the sequel we choose $u = (u_n)_{n \in \mathbb{N}}$ with $u_n = 0$ for all $n \in \mathbb{N}$ as in [29] or in [8]. Indeed, this choice has no effect on the asymptotic behavior of estimators.

Now, for any number of periods $K \geq 1$, any instants of breaks $\underline{t} \in \mathcal{F}_K$ and any parameters on each periods $\underline{\theta} \in \Theta^K$, the global QLIK contrast $\widehat{J}_n$ is given by the expression :

$$(QLIK) \quad \widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2 \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k).$$

Since K^* has to be estimated, define the QLIK contrast penalized by the number of periods, called *penQLIK* contrast, by

$$(penQLIK) \quad \widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n \, K \tag{2.4}$$

where $\kappa_n \leq n$ is called the regularization parameter and will be fixed later. Suppose that an upper bound $K_{max} > 0$ of the number of periods is known. Our estimator is defined as one of the minimizers of the penalized contrast :

$$(\widehat{K}_n, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) \in \underset{1 \leq K \leq K_{\max}}{\text{Argmin}} \; \underset{(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K}{\text{Argmin}} \; (\widetilde{J}_n(K, \underline{t}, \underline{\theta})) \quad \text{and} \quad \widehat{\underline{\tau}}_n = \frac{\widehat{\underline{t}}_n}{n}. \tag{2.5}$$

As very often in model selection problems, the whole estimation procedure deeply depends on the choice of the regularization parameters (κ_n). To be robust in possible dependence over distinct periods, the regularization parameters (κ_n) have to be carefully chosen :

Assumption H$_i$ ($i = 0, 1, 2$) : *For $0 \leq p \leq i$, the assumptions $A_p(f_\theta, \Theta)$, $A_p(M_\theta, \Theta)$*

(or respectively $A_p(h_\theta, \Theta)$) hold and $\theta_j^* \in \Theta$ for all $j = 1, \ldots, K^*$, Θ satisfying A for some $r \geq 1$. Denoting $c^* > 0$ a real number satisfying

$$c^* = \min_{j=1,\cdots,K^*} \big(-\log(\beta^{(0)}(\theta_j^*))/8\big) \quad \textit{or resp.} \quad \min_{j=1,\cdots,K^*} \big(-\log(\tilde{\beta}^{(0)}(\theta_j^*))/8\big)$$

the regularization parameters (κ_n) used in (2.4) satisfy $\kappa_n \wedge n\,\kappa_n^{-1} \to \infty$ with $n \to \infty$ and for all $j = 1, \cdots K^*$:

$$\sum_{k\geq 2} \kappa_k^{-(r/4\wedge 1)} \Big(\sum_{\ell \geq kc^*/\log(k)} \beta_\ell^{(p)}(\Theta)\Big)^{(r/4\wedge 1)} < \infty \quad \textit{or resp.} \quad \sum_{k\geq 2} \kappa_k^{-(r/4\wedge 1)} \Big(\sum_{\ell \geq kc^*/\log(k)} \tilde{\beta}_\ell^{(p)}(\Theta)\Big)^{(r/4\wedge 1)} < \infty \qquad (2.6)$$

The assumption H_i is interesting as it links the decrease rate of the Lipschitz coefficients and the penalty term of (2.4). The classical BIC corresponds to regularization parameters of the order of $\log(n)$. This choice is possible if the Lipschitz coefficients decrease exponentially fast, which hold for all models $\mathrm{M}(f_\theta, M_\theta)$ with finite order (see below). However, if the decrease of the Lipschitz coefficients is polynomial only regularization parameters satisfying $\kappa_n >> \log(n)$ satisfy H_i. Moreover, whatever the decay of the Lipschitzian coefficients, the estimation is more robust (with respect to the dependence over distinct segments) for the largest regularization parameter. More precisely, consider the following two paradigmatic examples for which (κ_n) satisfies conditions (2.6) (also used in [47]) :

(1) geometric case : if $\alpha_\ell^{(i)}(f_\theta, \Theta) + \alpha_\ell^{(i)}(M_\theta, \Theta) + \alpha_\ell^{(i)}(h_\theta, \Theta) = O(a^\ell)$ with $0 \leq a < 1$, then any choice of regularization parameters (κ_n) such that $\kappa_n \to \infty$ and $\kappa_n = o(n)$, satisfy (2.6) (for instance κ_n of order $\log n$ as in the BIC or MDL approach).

(2) Riemanian case : if $\alpha_\ell^{(i)}(f_\theta, \Theta) + \alpha_\ell^{(i)}(M_\theta, \Theta) + \alpha_\ell^{(i)}(h_\theta, \Theta) = O(\ell^{-\gamma})$ with $\gamma > 1$,
- if $\gamma > 1 + (1 \vee 4r^{-1})$, then any choice of (κ_n) such that $\kappa_n \to \infty$ and $\kappa_n = o(n)$ satisfy (2.6).
- if $(1 \vee 4r^{-1}) < \gamma \leq 1 + (1 \vee 4r^{-1})$, then any choice of (κ_n) such that $O(\kappa_n) = n^{1-\gamma+(1\vee 4r^{-1})}(\log n)^\delta$ with $\delta > \gamma - 1 + (1 \vee 4r^{-1})$ and $\kappa_n = o(n)$ can be chosen. However any of these choices satisfy $\kappa_n >> \log n$.

Remark 2.3.2. The sequence (δ_n) with $\delta_n := nc^*/\log n$ appearing in (2.6) is the size of "small" blocks that are excluded from the original observations to deal with the possible dependence between period. It is the theoretically size below which we do not distinguish the breaks due to the dependence. This size depends on the real model and is unknown.

2.3.2 Consistency of $(\widehat{K}_n, \underline{\hat{t}}_n, \underline{\widehat{\theta}}_n)$

For establishing the consistency, we add the couple of following classical assumptions in the problem of break detection :

Assumption B : $\displaystyle\min_{j=1,\cdots,K^*-1} \|\theta_{j+1}^* - \theta_j^*\| > 0.$

Furthermore, the distance between instants of breaks cannot be too small :

Assumption C : *there exists a vector $\underline{\tau}^* = (\tau_1^*, \ldots, \tau_{K^*-1}^*)$ with $0 < \tau_1^* < \ldots < \tau_{K^*-1}^* < 1$ called the vector of breaks such that for $j = 1, \cdots, K^*$, $t_j^* = [n\tau_j^*]$ (where $[x]$ is the floor of x). The is called the vector of breaks.*

Even if the length of T_j^* has asymptotically the same order than n, the dependence with respect to n of t_j^*, t_k, T_j^* and T_k are omitted for notational convenience.

Remark 2.3.3. The assumption C implies that the length of each segment tends to infinity at the same rate as n. We will introduce a size $u_n << n$ which represents the lower bound on the accuracy of the approximation of the lengths of the segments. This minimum size is needed for the numerical computation of the criteria. For the ARMA and GARCH model, $u_n = \mathcal{O}((\log n)^\delta)$ can be chosen for $1 \leq \delta \leq 2$.

We are now ready to prove the consistency of the penalized QLIK contrast :

Theorem 2.3.1. Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_0 are satisfied with $r \geq 2$. If $K_{\max} \geq K^*$ then :

$$(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (K^*, \underline{\tau}^*, \underline{\theta}^*).$$

Note that if K^* is known, we can relax the assumptions for the consistency by taking $\kappa_n = 1$ for all n as the penalty term in (2.4) does not matter. If K^* is unknown and $r = 2$, then a robust choice to any geometric or Riemanian dependence is $\kappa_n \propto n/\log n$. However, such large regularization parameters always over-penalized in practice.

2.3.3 Rates of convergence of the estimators

To state the rates of convergence of the estimators $\widehat{\underline{\tau}}_n$ and $\widehat{\underline{\theta}}_n$, we need to work under stronger moment and regularity assumptions. By convention, if the vectors $\widehat{\underline{t}}_n$ and $\underline{t}^*$ do not have the same length, complete the shorter of the 2 vectors with n before computing the norm $\|\widehat{\underline{t}}_n - \underline{t}^*\|_m$.

Theorem 2.3.2. Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_2 are satisfied with $r \geq 4$. If $K_{\max} \geq K^*$ then the sequence $(\|\widehat{\underline{t}}_n - \underline{t}^*\|_m)_{n>1}$ is uniformly tight in probability, *i.e.*

$$\lim_{\delta\to\infty} \lim_{n\to\infty} \mathbb{P}(\|\widehat{\underline{t}}_n - \underline{t}^*\|_m > \delta) = 0. \tag{2.7}$$

This theorem induces that $w_n^{-1} \|\widehat{\underline{t}}_n - \underline{t}^*\|_m \xrightarrow{P} 0$ for any sequence $(w_n)_n$ such that $w_n \to \infty$ and therefore $\|\widehat{\underline{t}}_n - \underline{t}^*\|_m = o_P(w_n)$: the convergence rate is arbitrary close to $O_P(1)$. This is the same convergence rate as in the case where $(X_t)_t$ is a sequence of independent r.v. (see for instance [7]). Such convergence rate was already reached in the frame of piecewise linear regression with innovations satisfying a mixing property in [51].

Let us turn now to the convergence rate of the estimator of parameters θ^*_j. By convention if $\widehat{K}_n < K^*$, set $\widehat{T}_j = \widehat{T}_{\widehat{K}_n}$ for $j \in \{\widehat{K}_n, \ldots, K^*\}$. Then,

Theorem 2.3.3. Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_2 are satisfied with $r \geq 4$ and $\kappa_n = O(\sqrt{n})$. Then if $\theta^*_j \in \overset{\circ}{\Theta}$ for all $j = 1, \cdots, K^*$, we have

$$\sqrt{n^*_j}\,(\widehat{\theta}_n(\widehat{T}_j) - \theta^*_j) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, F(\theta^*_j)^{-1} G(\theta^*_j) F(\theta^*_j)^{-1}), \tag{2.8}$$

where, using $q_{0,j}$ defined in (2.11), the matrix F and G are such that

$$(F(\theta^*_j))_{k,l} = I\!E\Big(\frac{\partial^2 q_{0,j}(\theta^*_j)}{\partial\theta_k \partial\theta_l}\Big) \text{ and } (G(\theta^*_j))_{k,l} = I\!E\Big(\frac{\partial q_{0,j}(\theta^*_j)}{\partial\theta_k}\frac{\partial q_{0,j}(\theta^*_j)}{\partial\theta_l}\Big). \tag{2.9}$$

In Theorem 2.3.3, a condition on the rate of convergence of κ_n is added. The most robust choice for the regularization parameter corresponds to $\kappa_n \propto \sqrt{n}$ as it corresponds to the most general problem (2.3) (see above). However, by assumption H_2 it excludes models with finite moments $r \geq 4$ satisfying : $\ell^{-\gamma} = O(\alpha^{(i)}_\ell(f_\theta, \Theta) + \alpha^{(i)}_\ell(M_\theta, \Theta))$ or $\ell^{-\gamma} = \alpha^{(i)}_\ell(h_\theta, \Theta))$ with $1 < \gamma \leq 3/2$ for some $i = 0, 1, 2$. For these models the consistency for $\widehat{\underline{\tau}}_n$ holds but we do not get any rate of convergence for $\widehat{\underline{\theta}}_n$.

2.4 Some examples

2.4.1 AR(∞) models

Consider $AR(\infty)$ with K^*-1 breaks defined by the equation :

$$X_t = \sum_{k\geq 1} \phi_k(\theta^*_j) X_{t-k} + \xi_t \ , \quad t^*_{j-1} < t \leq t^*_j \ , \ j = 1, \cdots, K^*.$$

This is the case of the problem (2.3) with models $\mathcal{M}_{T^*_i}(f_\theta, M_\theta)$ where $f_\theta(x_1, \cdots) = \sum_{k\geq 1} \phi_k(\theta) x_k$ and $M_\theta \equiv 1$. Assume that Θ is a compact set such that $\sum_{k\geq 1} \|\phi_k(\theta)\|_\Theta < 1$. Thus $\Theta = \Theta$ for any $r \geq 1$ satisfying $I\!E|\xi_0|^r < \infty$. Then Assumptions D(Θ) and $A_0(f_\theta, \Theta)$ hold automatically with $\underline{h} = 1$ and $\alpha^{(0)}_k(f_\theta, \Theta) = \|\phi_k(\theta)\|_\Theta$. Then,

- Assume that Id(Θ) holds and that there exists $r \geq 2$ such that $I\!E|\xi_0|^r < \infty$. If there exists $\gamma > 1 \vee 4r^{-1}$ such that $\|\phi_k(\theta)\|_\Theta = O(k^{-\gamma})$ for all $k \geq 1$, the choice $\kappa_n = n/\log n$ ensures the strong consistency of $(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n)$.
- Moreover, if $I\!E|\xi_0|^4 < \infty$, $\gamma > 3/2$ and ϕ_k twice differentiable satisfying $\|\phi'_k(\theta)\|_\Theta = O(k^{-\gamma})$ and $\|\phi''_k(\theta)\|_\Theta = O(k^{-\gamma})$, the choice $\kappa_n = \sqrt{n}$ ensures the convergence (2.7) of $\widehat{\underline{t}}_n$ and the CLT (2.8) satisfied by $\widehat{\underline{\theta}}_n(\widehat{T}_j)$ for all j.

Note that this problem of change detection was considered by Davis *et al.* in [24] but moments $r > 4$ are required. In Davis *et al.* [25], the same problem for another break model for AR processes is studied. However, in both these papers, the process is supposed to be independent from one block to another and stationary on each block.

2.4.2 ARCH(∞) models

Consider an $ARCH(\infty)$ model with K^*-1 breaks defined by :

$$X_t = \Big(\psi_0(\theta^*_j) + \sum_{k=1}^{\infty} \psi_k(\theta^*_j) X^2_{t-k}\Big)^{1/2} \xi_t \ , \quad t^*_{j-1} < t \leq t^*_j \ , \ j = 1, \cdots, K^*,$$

where $(\psi_k(\theta))_{k\geq 0}$ is a sequence of positive real numbers and $I\!E(\xi_0^2) = 1$. Note that $h_\theta((x_k)_{k\in\mathbb{N}}) = \psi_0(\theta) + \sum_{k=1}^\infty \psi_k(\theta) x_k^2$ and $f_\theta = 0$. Assume that Θ is a compact set such that $\sum_{k\geq 1} \|\psi_k(\theta)\|_\Theta < 1$, then $\Theta(2) = \Theta$. Assume that $\inf_{\theta\in\Theta} \psi_0(\theta) > 0$ which ensures that $D(\Theta)$ and $Id(\Theta)$ hold.

- If there exists $\gamma > 2$ such that $\|\psi_k(\theta)\|_\Theta = O(k^{-\gamma})$ for all $k \geq 1$, then the choice of $\kappa_n = n/\log n$ leads to the consistency of $(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n)$ when $\theta^*_j \in \Theta$ for all j.
- Moreover, if $I\!E|\xi_0|^4 < \infty$, $\Theta(4)$ is a compact set such s $\theta^*_j \in \overset{\circ}{\Theta}(4)$ for all j, and if ψ_k is a twice differentiable function satisfying $\|\psi'_k(\theta)\|_\Theta = O(k^{-\gamma})$ and $\|\psi''_k(\theta)\|_\Theta = O(k^{-\gamma})$ with $\gamma > 3/2$, then the choice of $\kappa_n = \sqrt{n}$ ensures the convergence (2.7) and the CLT (2.8) satisfied by $\widehat{\underline{\theta}}_n(\widehat{T}_j)$ for all j.

This problem of break detection was already studied by Kokoszka and Leipus in [45] but they obtained the consistency of their procedure under stronger assumptions.

Example 1. Let us detail the GARCH(p,q) model with K^*-1 breaks defined by :

$$X_t = \sigma_t \xi_t \ , \ \sigma_t^2 = a^*_{0,j} + \sum_{k=1}^{q} a^*_{k,j} X^2_{t-k} + \sum_{k=1}^{p} b^*_{k,j} \sigma^2_{t-k} \quad t^*_{j-1} < t \leq t^*_j \ , \ j = 1, \cdots, K^*$$

with $I\!E(\xi_0^2) = 1$. Assume that for any $\theta = (a_0, \ldots, a_q, b_1, \ldots, b_p) \in \Theta$ then $a_k \geq 0$, $b_k \geq 0$ and $\sum_{k=1}^p b_k < 1$. Then, there exists (see Nelson and Cao [60]) a nonnegative sequence $(\psi_k(\theta))_k$ such that $\sigma_t^2 = \psi_0(\theta) + \sum_{k\geq 1} \psi_k(\theta) X_{t-k}^2$. Remark that this sequence is twice differentiable with respect to θ and that its derivatives are exponentially decreasing. Moreover for any $\theta \in \Theta$, $\sum_{k\geq 1} \psi_k(\theta) \leq \big(\sum_{k=1}^q a_k\big)/\big(1-\sum_{k=1}^p b_k\big)$ and

$$\Theta = \left\{ \theta \in \Theta \ , \ (I\!E|\xi_0|^r)^{2/r} \sum_{k=1}^{q} a_k + \sum_{k=1}^{p} b_k < 1 \right\}.$$

Then if $\sum_{k=1}^q a_{k,j}^* + \sum_{k=1}^p b_{k,j}^* < 1$ for all j (case $r \geq 2$), our estimation procedure associated with a regularization parameter $\kappa_n K$ for any $1 << \kappa_n << n$ is consistent. Moreover, if $(I\!E|\xi_0|^4)^{1/2} \sum_{k=1}^q a_{k,j}^* + \sum_{k=1}^p b_{k,j}^* < 1$ for all j, then our procedure with a regularization parameter $1 << \kappa_n = O(\sqrt{n})$ allows the same rates of convergence than in the case where (X_t) are independent random variables. For example, a BIC-type regularization parameter $\kappa_n \propto \log n$ as in [25] can be chosen in this case.

2.4.3 Estimates breaks in TARCH(∞) model

Consider a TARCH(∞) model with breaks defined by :

$$X_t = \sigma_t \, \xi_t \ , \ \sigma_t = b_0(\theta_j^*) + \sum_{k\geq 1} \Big(b_k^+(\theta_j^*) \max(X_{t-k}, 0) - b_k^-(\theta_j^*) \min(X_{t-k}, 0) \Big),$$

with $t_{j-1}^* < t \leq t_j^*$, $j = 1, \cdots, K^*$, where $(b_k^+)_{k\geq 0}$, $(b_k^-)_{k\geq 0}$ are sequence of positive real numbers satisfying $\|b_0(\theta)\|_\Theta > 0$ and $\sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < \infty$. Then $f_\theta = 0$ and $(A_0(M_\theta, \Theta))$ holds with $\alpha_k^{(0)}(M_\theta, \Theta) = \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta)$

- Assume $\|\xi_0\|_r^{1/r} \sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < 1$ for $r \geq 2$. If there exists $\gamma > 1 \vee 4r^{-1}$ such that $\max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) = O(k^{-\gamma})$ for all $k \geq 1$, then $\kappa_n = n/\log n$ leads to the consistency of $(\widehat{K}_n, \underline{\widehat{\tau}}_n, \underline{\widehat{\theta}}_n)$ when $\theta_j^* \in \Theta(2)$ for all j.
- Moreover, if $I\!E|\xi_0|^4 < \infty$ and b_k^+, b_k^- are twice differentiable satisfying $\|\partial b_k^+(\theta)/\partial\theta\|_\Theta = O(k^{-\gamma})$ and $\|\partial^2 b_k^+(\theta)/\partial\theta^2\|_\Theta = O(k^{-\gamma})$ with $\gamma > 3/2$ (the same for b_k^-), then $\kappa_n = \sqrt{n}$ ensures the convergence (2.7) and the CLT (2.8) satisfied by $\underline{\widehat{\theta}}_n(\widehat{T}_j)$ for all j (with $\theta_j^* \in \overset{\circ}{\Theta}(4)$).

To our knowledge, these results are the first one concerning the change detection for TARCH(∞).

2.5 Some simulations results

The procedure is implemented on the R software (developed by the CRAN project). Since we proceed with not so large samples ($n \leq 2000$), the consistency of $\widehat{K}_n$ is often not obtained for the most robust theoretical choice of $\kappa_n = \sqrt{n}$. As a consequence, for numerical applications, we chose a data-driven procedure for computing the regularization parameter κ_n. Thus, κ_n is calibrated using the slope estimation procedure of Baudry et al. [12]. Once obtained the regularization parameter κ_n, the dynamic programming algorithm (see [39]) is used to minimize the criteria. Remark that we could also use the genetic algorithm and the approximated likelihood of [25] to speed up the procedure.

2.5.1 The slope estimation procedure

The heuristic of the procedure is that the criteria (here QLIK) is a linear transformation of the penalties (here the number of periods K) for the most complex models (with

K close to K_{max}). This slope should be close to $-\kappa_n/2$. This procedure has already been used in [12] for breaks detection in an i.i.d. context. We adapt it to the case of dependence (details are omitted for the common part with the iid case and we refer the interested reader to [12]).

By construction, the procedure is very sensitive to the choice of K_{max} as only complex models are used to estimate the slope. As discussed in the Remark 2.3.3, we only consider periods of length larger than u_n and we can a priori fix K_{max} smaller than $[n/u_n]$. Therefore, the slope estimation procedure consider only the linear part of $-QLIK$ with $K \leq K_{max}$. The concrete procedure (see examples below) is :

1. For each $1 \leq K \leq K_{max}$, draw $(K, -\min_{\underline{t},\underline{\theta}} QLIK(K))_{1\leq K\leq K_{max}}$. Then compute the slope of the linear part : this slope is $\hat{\kappa}_n/2$.
2. Using $\kappa_n = \hat{\kappa}_n$, draw $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))_{1\leq K\leq K_{max}}$. This curve has a global minimum at $\widehat{K}_n$.

2.5.2 Implementation details

We assume that the regularization parameter is known (for instance $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$ or $\kappa_n = \sqrt{n}$). In this section, we give more details on how to compute $\widehat{K}_n$ and the optimal configuration of the breaks by using the dynamic programming algorithm. The basic idea of this algorithm is that : for a given $1 \leq K \leq K_{max}$, if $(t_1, \cdots, t_{K-1}, t)$ is an optimal configuration of $X_1, \cdots, X_t$ into K segments, then $(t_1, \cdots, t_{K-1})$ is an optimal configuration of $X_1, \cdots, X_{t_{K-1}}$ into $K-1$ segments.
For $1 \leq i \leq l \leq n$, denote $T_{i,l} = \{i, i+1, \cdots, l\}$ and let ML be the upper triangular matrix of dimension $n \times n$ with $ML_{i,l} = \widehat{L}(T_{i,l}, \widehat{\theta}_n(T_{i,l}))$ for $i \leq l$. The estimated number of segment $\widehat{K}_n$ and the corresponding optimal configuration can be obtained as follow :

1. Let C be an upper triangular matrix of dimension $K_{max} \times n$. For $1 \leq K \leq K_{max}$ and $K \leq t \leq n$, $C_{K,t}$ will be the minimum penalized criteria of $X_1, \cdots, X_t$ into K segments. Therefore, for $t = 1, \cdots, n$ $C_{1,t} = -2ML_{1,t} + \kappa_n$ and the relation $C_{K+1,t} = \min_{K\leq l\leq t-1} (C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied. Hence, $\widehat{K}_n = \underset{K\leq l\leq K_{max}}{\text{Argmin}} (C_{K,n})$.
2. Let Z be an upper triangular matrix of dimension $(K_{max}-1) \times n$. For $1 \leq K \leq (K_{max}-1)$ and $K+1 \leq t \leq n$, $Z_{K,t}$ will be the Kth potential break-point of $X_1, \cdots, X_t$. Therefore, the relation $Z_{K,t} = \underset{K\leq l\leq t-1}{\text{Argmin}} (C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied for $K = 1, \cdots, K_{max}-1$ and the break-point are obtained as follow : set $\widehat{t}_{\widehat{K}_n} = n$, $\widehat{t}_1 = 1$ and for $K = \widehat{K}_n - 1, \cdots, 2$, $\widehat{t}_K = Z_{K,\widehat{t}_{K+1}}$.

Note that the above procedure requires $\mathcal{O}(n^2)$ operations, instead of $\mathcal{O}(n^{K_{max}})$ if the standard procedure is used.

Remark 2.5.1. The minimum description length (MDL in the sequel) criterion (see [25]) is defined in our setting by :

$$MDL(K, \underline{t}, \underline{\theta}) := \log_+(K-1) + K \log n + \frac{d}{2} \sum_{k=1}^{K} \log n_k - \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k)$$

where $\log_+(x) = 0$ if $x \leq 1$ and $\log_+(x) = \log x$ if $x > 1$. We can also write :

$$MDL(K, \underline{t}, \underline{\theta}) = \widetilde{MDL}(K, \underline{t}, \underline{\theta}) + K \log n$$

where

$$\widetilde{MDL}(K, \underline{t}, \underline{\theta}) = -\sum_{k=1}^{K} \Big(\widehat{L}_n(T_k, \theta_k) - \frac{d}{2} \log n_k - \log_+(k-1) + \log_+(k-2) \Big).$$

Hence, the MDL criterion can be seen as a penalized criterion and the dynamic programming algorithm described above can be used to find the optimal configuration.

2.5.3 Results of simulations

AR(1) models : we consider the problem (2.3) for a AR(1) :

$$X_t = \theta_j^* X_{t-1} + \xi_t \qquad \forall t \in T_j, \quad \forall j \in \{1, \cdots, K^*\}.$$

For $n = 500$ and $n = 1000$, we generate a sample $(X_1, \cdots, X_n)$ in the following situations :

- **scenario $\mathbf{A}_0$** : $\theta^*(1) = 0.5$ is constant $(K^* = 1)$;
- **scenario $\mathbf{A}_1$** : $\theta^*(1) = 0.5$ changes to $\theta^*(2) = 0.2$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario $\mathbf{A}_2$** : $\theta^*(1) = 0.7$ changes to $\theta^*(2) = 0.9$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario $\mathbf{A}_3$** : $\theta^*(1) = 0.5$ changes to $\theta^*(2) = 0.3$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = 0.7$ at $t_2^* = 0.7n$ $(K^* = 3)$;
- **scenario $\mathbf{A}_4$** : $\theta^*(1) = 0.7$ changes to $\theta^*(2) = 0.9$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = 0.6$ at $t_2^* = 0.7n$ $(K^* = 3)$.

The regularization parameter is chosen by using the slope estimation presented above (Subsection 2.5.1). Figure 2.1 represents the slope of the linear part of the $-QLIK$ criteria (minimized in $(\underline{t}, \underline{\theta})$) in scenario A_4 for $n = 500$ and $n = 1000$. Thus, by referring to the

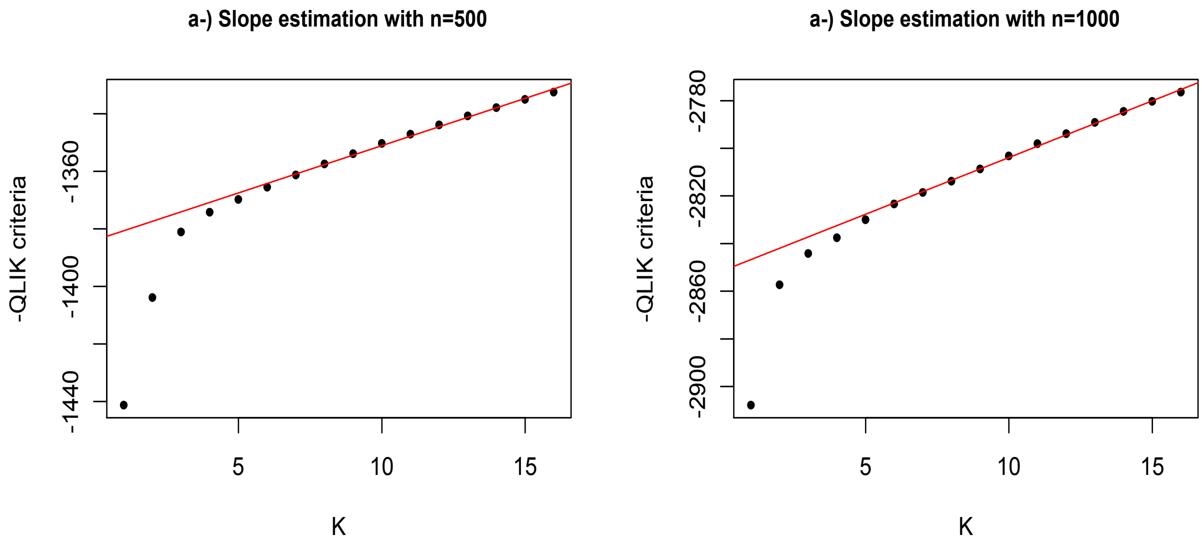

FIGURE 2.1 – The curve of $-\min_{t,\theta} QLIK$ for $1 \leq K \leq K_{max}$ for AR(1) process in scenario A_4. The solid line represents the linear part of this curve with slope $\hat{\kappa}_n/2 = 3.47$ when $n = 500$ and $\hat{\kappa}_n/2 = 4.90$ when $n = 1000$

Figure 2.1 we obtain $\hat{\kappa}_n \approx 7.0$ for $n = 500$ and $\hat{\kappa}_n \approx 9.8$ for $n = 1000$.
We are going to minimize the $penQLIK$ in $(K, \underline{t}, \underline{\theta})$, with $1 \leq K \leq K_{max}$ and $\kappa_n = \hat{\kappa}_n$. Figure 2.2 represents the points $(K, \min_{\underline{t}, \underline{\theta}} penQLIK(K))$ for $1 \leq K \leq K_{max} = 10$.

One can easily read on the Figure 2.2, the estimated values $\widehat{K}_n = 4$ for $n = 500$ and

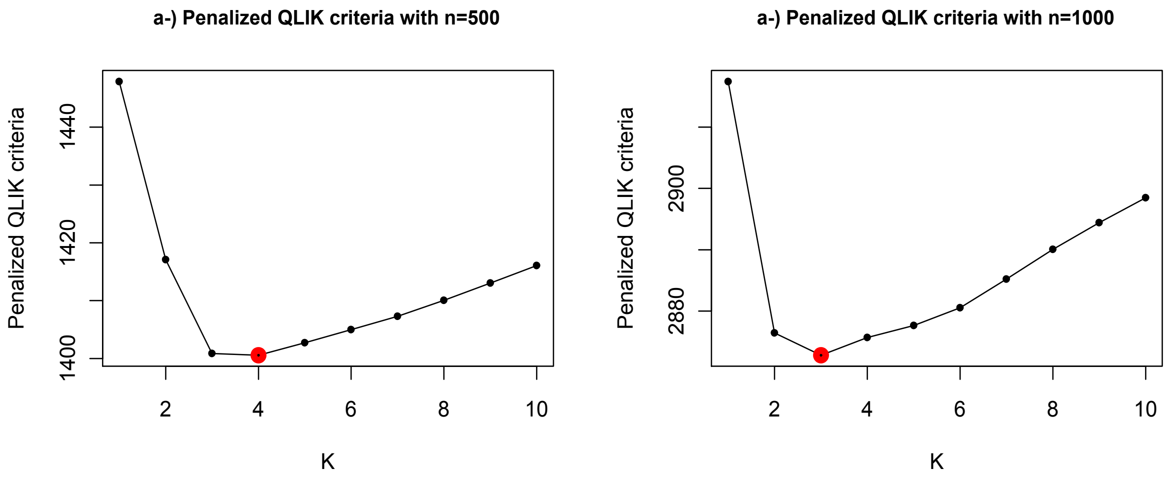

Figure 2.2 – The graph$(K, \min_{t,\theta} penQLIK(K))$ for $1 \leq K \leq K_{max} = 10$ for AR(1) in scenario A_4.

$\widehat{K}_n = 3$ for $n = 1000$ (the estimated number of break is $\widehat{K}_n - 1$). Moreover, the estimated instants of break are $\widehat{\underline{t}}_n = (146, 228, 357)$ ($\underline{t}^* = (150, 350)$) for $n = 500$ and $\widehat{\underline{t}}_n = (282, 687)$ ($\underline{t}^* = (300, 700)$) for $n = 1000$. Figure 2.3 shows the estimated break points for two trajectories ($n = 500$ and $n = 1000$) for AR(1) processes following scenario $\mathbf{A}_4$ with two changes.

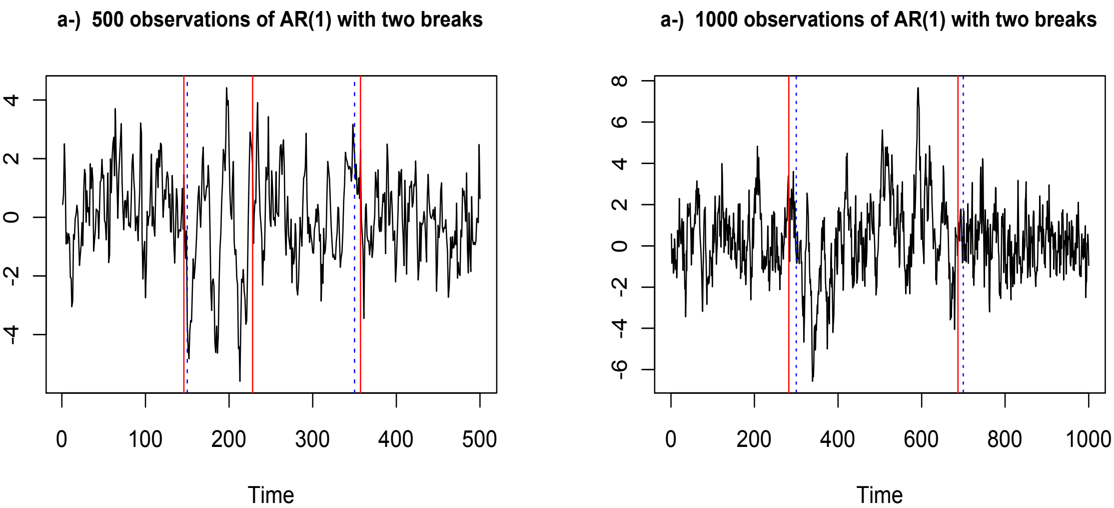

Figure 2.3 – The estimated of breakpoints for a trajectory of AR(1) process in scenario A_4. The solid lines represent the estimated break instants and the dotted lines represent the true ones.

Now, 100 independent replications of a AR(1) process are generated following the scenarios $\mathbf{A}_0$-$\mathbf{A}_4$. For each replication, the estimated number of segments is computed using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and using MDL procedure and Table 2.1

indicates the proportions of number of replications (frequencies) where the true number of breaks is achieved following the scenarios $\mathbf{A}_0$-$\mathbf{A}_4$.
For the replications of scenario $\mathbf{A}_4$, where the true number of break is fitted ($\widehat{K}_n = 3$), the average of the estimated parameters are computed and shown in Table 2.2.

AR(2) models : we consider the problem (2.3) for a AR(2) :

$$X_t = \phi_1^*(j)\, X_{t-1} + \phi_2^*(j)\, X_{t-2} + \xi_t \qquad \forall t \in T_j, \;\; \forall j \in \{1, \cdots, K^*\}.$$

Denote $\theta^*(j) = (\phi_1^*(j), \phi_2^*(j))$. For $n = 500$ and $n = 1000$, we generate a sample $(X_1, \cdots, X_n)$ with one break at $t^* = 0.5n$ in the following situations :

- **scenario $\mathbf{B}_0$** : $\theta^*(1) = (0.4, 0.3)$ is constant ($K^* = 1$) ;
- **scenario $\mathbf{B}_1$** : $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.1, 0.3)$;
- **scenario $\mathbf{B}_2$** : $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.2, 0.5)$;
- **scenario $\mathbf{B}_3$** : $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.6, 0.1)$.

100 independent replications of a AR(2) process are generated following the scenarios $\mathbf{B}_0$-$\mathbf{B}_3$. It is evaluated the performance of the procedure using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and the one of MDL procedure. Table 2.3 indicates the proportions of number of replications (frequencies) where the true number of breaks is achieved following the scenarios $\mathbf{B}_0$-$\mathbf{B}_3$.

GARCH(1,1) models : we consider examples of problem (2.3) when X is a $GARCH(1,1)$ process on each period :

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = a_0^*(j) + a_1^*(j) X_t^2 + b_1^*(j)\sigma_t^2 \qquad \forall t \in T_j^*, \;\; \forall j \in \{1, \cdots, K^*\}. \quad (2.10)$$

Thus $\theta^* = (a_0^*, a_1^*, b_1^*)$. For $n = 500$ and $n = 1000$, we generate $(X_1, \cdots, X_n)$ in the following situation :

- **scenario $\mathbf{G}_0$** : $\theta^*(1) = (0.5, 0.2, 0.2)$ is constant ($K^* = 1$) ;
- **scenario $\mathbf{G}_1$** : $\theta^*(1) = (0.5, 0.2, 0.2)$ changes to $\theta^*(2) = (0.5, 0.2, 0.6)$ at $t^* = 0.5n$ ($K^* = 2$) ;
- **scenario $\mathbf{G}_2$** : $\theta^*(1) = (0.5, 0.6, 0.2)$ changes to $\theta^*(2) = (1, 0.6, 0.2)$ at $t^* = 0.5n$ ($K^* = 2$) ;
- **scenario $\mathbf{G}_3$** : $\theta^*(1) = (0.5, 0.2, 0.2)$ changes to $\theta^*(2) = (0.5, 0.2, 0.0)$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = (0.1, 0.2, 0.0)$ at $t_2^* = 0.7n$ ($K^* = 3$) ;
- **scenario $\mathbf{G}_4$** : $\theta^*(1) = (0.5, 0.6, 0.2)$ changes to $\theta^*(2) = (1, 0.6, 0.2)$ (at $t_1^* = 0.3n$) which changes to $\theta^*(3) = (1, 0.2, 0.2)$ at $t_2^* = 0.7n$ ($K^* = 3$).

Figure 2.4 shows an example of scenario $\mathbf{G}_4$ where one break is fitted with $\hat{\kappa}_n \approx 12.7$ for $n = 500$ and two breaks with $\hat{\kappa}_n \approx 18.3$ for $n = 1000$; we obtain, $\widehat{\underline{t}}_n = 168$ (while $\underline{t}^* = (150, 350)$) for $n = 500$ and $\widehat{\underline{t}}_n = (307, 725)$ (while $\underline{t}^* = (300, 700)$) for $n = 1000$.

Now, 100 independent replications of GARCH$(1,1)$ processes are generated following the scenarios $\mathbf{G}_0$-$\mathbf{G}_4$. For each replication, the estimated number of segment is computed using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and using MDL procedure and Table 2.4 indicates the proportions of replications (frequencies) when the true number of breaks is achieved following the scenarios $\mathbf{G}_0$-$\mathbf{G}_4$.

For the replications of the scenario $\mathbf{G}_2$, when the true number of break is fitted ($\widehat{K}_n = 2 = K^*$), the average of the estimated parameters are computed and shown in Table 2.5.

Model			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$
scenario A_0	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.74	0.00	0.26
$(K^* = 1)$		$\kappa_n = \log n$	0.50	0.00	0.50
		$\kappa_n = \sqrt{n}$	0.94	0.00	0.06
		MDL procedure	0.95	0.00	0.05
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.81	0.00	0.20
		$\kappa_n = \log n$	0.43	0.00	0.57
		$\kappa_n = \sqrt{n}$	1.00	0.00	0.00
		MDL procedure	0.97	0.00	0.03
scenario A_1	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.52	0.06	0.42
$(K^* = 2)$		$\kappa_n = \log n$	0.40	0.04	0.56
		$\kappa_n = \sqrt{n}$	0.23	0.77	0.00
		MDL procedure	0.44	0.56	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.78	0.00	0.22
		$\kappa_n = \log n$	0.40	0.00	0.60
		$\kappa_n = \sqrt{n}$	0.38	0.62	0.00
		MDL procedure	0.87	0.13	0.00
scenario A_2	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.48	0.00	0.52
$(K^* = 2)$		$\kappa_n = \log n$	0.17	0.00	0.83
		$\kappa_n = \sqrt{n}$	0.29	0.71	0.00
		MDL procedure	0.56	0.44	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.76	0.00	0.24
		$\kappa_n = \log n$	0.06	0.00	0.94
		$\kappa_n = \sqrt{n}$	0.57	0.43	0.00
		MDL procedure	0.89	0.07	0.04
scenario A_3	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.45	0.32	0.23
$(K^* = 3)$		$\kappa_n = \log n$	0.37	0.26	0.37
		$\kappa_n = \sqrt{n}$	0.00	1.00	0.00
		MDL procedure	0.01	0.99	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.61	0.13	0.26
		$\kappa_n = \log n$	0.39	0.00	0.61
		$\kappa_n = \sqrt{n}$	0.00	1.00	0.00
		MDL procedure	0.20	0.80	0.00
scenario A_4	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.53	0.12	0.35
$(K^* = 3)$		$\kappa_n = \log n$	0.28	0.06	0.66
		$\kappa_n = \sqrt{n}$	0.04	0.96	0.00
		MDL procedure	0.09	0.91	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.75	0.00	0.25
		$\kappa_n = \log n$	0.12	0.00	0.88
		$\kappa_n = \sqrt{n}$	0.06	0.94	0.00
		MDL procedure	0.54	0.46	0.00

Table 2.1 – Frequencies of the number of breaks estimated after 100 replications for AR(1) process following scenarios $\mathbf{A}_0$-$\mathbf{A}_4$.

n		Mean ± standard deviation $\widehat{\tau}_1$	$\widehat{\tau}_2$	Mean of $\Vert\widehat{\underline{\tau}}_n - \underline{\tau}^*\Vert$	Mean of $\widehat{\theta}(j)$, $j = 1, 2, 3$
$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.310 ± 0.078	0.719 ± 0.049	0.070	0.668 ; 0.894 ; 0.571
	$\kappa_n = \log n$	0.308 ± 0.064	0.718 ± 0.049	0.065	0.666 ; 0.898 ; 0.567
	$\kappa_n = \sqrt{n}$	0.323 ± 0.037	0.678 ± 0.016	0.045	0.600 ; 0.935 ; 0.561
	MDL procedure	0.316 ± 0.044	0.680 ± 0.013	0.042	0.637 ; 0.926 ; 0.577
$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.297 ± 0.078	0.691 ± 0.025	0.063	0.694 ; 0.894 ; 0.613
	$\kappa_n = \log n$	0.317 ± 0.038	0.710 ± 0.033	0.045	0.708 ; 0.874 ; 0.598
	$\kappa_n = \sqrt{n}$	0.341 ± 0.078	0.714 ± 0.023	0.046	0.640 ; 0.905 ; 0.528
	MDL Procedure	0.340 ± 0.085	0.702 ± 0.029	0.062	0.676 ; 0.911 ; 0.586

Table 2.2 – The estimated parameters for the replications of AR(1) processes following $\mathbf{A}_4$ satisfying $\widehat{K}_n = 3 = K^*$ (two changes estimated).

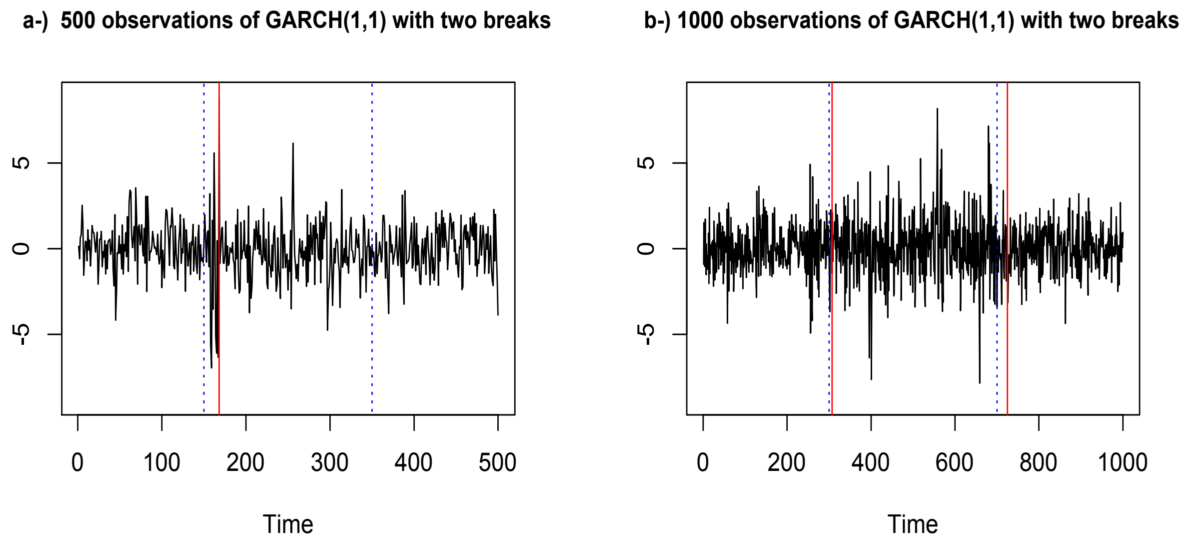

Figure 2.4 – A GARCH(1,1) process with 2 breaks ($K^* = 3$) following the scenario $\mathbf{G}_4$. The solid lines represent the estimated break instants and the dotted lines represent the true ones.

Finally, recall that in [25], the process is stationary on each segment and assumed to be independent from a segment to another. Davis *et al.* (2008) used the genetic algorithm to approximate the optimal values of the MDL criteria. We consider three of their scenarios with $n = 1000$ for GARCH$(1, 1)$ processes :

- **scenario A** : $\theta^*(1) = (0.4, 0.1, 0.5)$ is constant ($K^* = 1$) ;
- **scenario C** : $\theta^*(1) = (0.4, 0.1, 0.5)$ changes to $\theta^*(2) = (0.4, 0.1, 0.6)$ at $t^* = 0.5n$ ($K^* = 2$) ;
- **scenario J** : $\theta^*(1) = (0.1, 0.1, 0.8)$ changes to $\theta^*(2) = (0.5, 0.1, 0.8)$ at $t^* = 0.5n$ ($K^* = 2$).

Table 2.6 shows the results obtained with our penQLIK method with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and the results of the MDL procedure (obtained after 500 replications) taken in

Model			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$
scenario B_0	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.61	0.00	0.39
$(K^* = 1)$		$\kappa_n = \log n$	0.08	0.00	0.92
		$\kappa_n = \sqrt{n}$	0.94	0.00	0.06
		MDL procedure	0.92	0.00	0.08
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.79	0.00	0.21
		$\kappa_n = \log n$	0.06	0.00	0.94
		$\kappa_n = \sqrt{n}$	0.98	0.00	0.02
		MDL procedure	0.97	0.00	0.03
scenario B_1	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.63	0.04	0.33
$(K^* = 2)$		$\kappa_n = \log n$	0.15	0.00	0.85
		$\kappa_n = \sqrt{n}$	0.37	0.63	0.00
		MDL procedure	0.38	0.62	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.83	0.00	0.17
		$\kappa_n = \log n$	0.03	0.00	0.97
		$\kappa_n = \sqrt{n}$	0.60	0.40	0.00
		MDL procedure	0.85	0.15	0.00
scenario B_2	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.57	0.15	0.28
$(K^* = 2)$		$\kappa_n = \log n$	0.09	0.01	0.90
		$\kappa_n = \sqrt{n}$	0.10	0.90	0.00
		MDL procedure	0.11	0.89	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.78	0.08	0.14
		$\kappa_n = \log n$	0.05	0.00	0.95
		$\kappa_n = \sqrt{n}$	0.17	0.83	0.00
		MDL procedure	0.24	0.76	0.00
scenario B_3	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.41	0.25	0.34
$(K^* = 3)$		$\kappa_n = \log n$	0.08	0.03	0.89
		$\kappa_n = \sqrt{n}$	0.07	0.93	0.00
		MDL procedure	0.08	0.92	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.75	0.08	0.17
		$\kappa_n = \log n$	0.03	0.00	0.97
		$\kappa_n = \sqrt{n}$	0.19	0.81	0.00
		MDL procedure	0.22	0.78	0.00

Table 2.3 – Frequencies of the number of breaks estimated after 100 replications for AR(2) process following scenarios $\mathbf{B}_0$-$\mathbf{B}_3$.

Model			$\widehat{K_n} = K^*$	$\widehat{K_n} < K^*$	$\widehat{K_n} > K^*$
scenario G_0	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.44	0.00	0.56
$(K^* = 1)$		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.58	0.00	0.42
		MDL procedure	0.51	0.00	0.49
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.60	0.00	0.40
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.75	0.00	0.25
		MDL procedure	0.63	0.00	0.37
scenario G_1	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.42	0.12	0.46
$(K^* = 2)$		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.52	0.05	0.00
		MDL procedure	0.55	0.35	0.10
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.65	0.00	0.35
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.74	0.10	0.00
		MDL procedure	0.67	0.09	0.24
scenario G_2	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.52	0.20	0.28
$(K^* = 2)$		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.39	0.44	0.17
		MDL procedure	0.44	0.40	0.16
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.56	0.10	0.34
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.42	0.48	0.10
		MDL procedure	0.57	0.31	0.12
scenario G_3	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.41	0.28	0.31
$(K^* = 3)$		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.40	0.60	0.00
		MDL procedure	0.53	0.39	0.08
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.70	0.26	0.04
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.43	0.57	0.00
		MDL procedure	0.59	0.37	0.04
scenario G_4	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.30	0.55	0.15
$(K^* = 3)$		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.08	0.90	0.02
		MDL procedure	0.16	0.77	0.07
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.53	0.29	0.18
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.10	0.90	0.00
		MDL procedure	0.27	0.66	0.07

Table 2.4 – Frequencies of the number of breaks estimated after 100 replications for GARCH$(1,1)$ processes following the scenarios $\mathbf{G}_0$-$\mathbf{G}_4$.

n		Mean ± standard deviation $\widehat{\tau}$	Mean of $\lvert\widehat{\tau}-\tau^*\rvert$	Mean of $\widehat{\theta}(j)$, $j=1,2$
$n=500$	$\kappa_n=\widehat{\kappa}_n$	0.428 ± 0.245	0.176	(0.489, 0.513, 0.202)
				(1.027, 0.560, 0.220)
	$\kappa_n=\log n$	NA	NA	NA
	$\kappa_n=\sqrt{n}$	0.297 ± 0.260	0.253	(0.374, 0.376, 0.234)
				(1.026, 0.606, 0.182)
	MDL procedure	0.372 ± 0.275	0.222	(0.439, 0.437, 0.203)
				(0.980, 0.579, 0.224)
$n=1000$	$\kappa_n=\widehat{\kappa}_n$	0.466 ± 0.130	0.058	(0.455, 0.528, 0.201)
				(1.038, 0.595, 0.186)
	$\kappa_n=\log n$	NA	NA	NA
	$\kappa_n=\sqrt{n}$	0.302 ± 0.266	0.215	(0.276, 0.395, 0.354)
				(1.033, 0.582, 0.182)
	MDL procedure	0.469 ± 0.133	0.059	(0.455, 0.528, 0.201)
				(1.052, 0.597, 0.181)

Table 2.5 – The estimated parameters for the replications of GARCH$(1,1)$ processes following the scenario G_2 and satisfying $\widehat{K}_n = 2 = K^*$ (one break fitted).

Table I of [25].

Conclusion of simulations for AR(1), AR(2) and GARCH$(1,1)$ processes : The results of QLIK criteria with $\widehat{\kappa}_n$ and $\sqrt{n}$ penalty show that the probability $\Pr(\widehat{K}_n = K^*)$ increases as n increases in all scenarios as it can be deduced from the theory. This is not the case for $\log n$ penalty (see for instance the scenario $\mathbf{A}_2$). Comparing the results of scenarios $\mathbf{A}_1$ and $\mathbf{A}_2$ (or scenarios $\mathbf{A}_3$ and $\mathbf{A}_4$), the BIC penalty ($\kappa_n = \log n$) under-penalizes the number of breaks when the process is sufficiently dependent on its own past. More dependent the process, larger the probability to fit the true number of breaks with $\sqrt{n}$ or $\widehat{\kappa}_n$ penalty (except in the case $\mathbf{G}_2$ for $\sqrt{n}$ penalty). However in the case of two breaks, the $\sqrt{n}$ penalty over-penalizes the number of breaks contrarily with $\widehat{\kappa}_n$ penalty which provides the best results as well for AR(1) as for GARCH$(1,1)$ processes.
For the scenarios $\mathbf{A}_1$-$\mathbf{A}_4$, the change in the parameter induces a change in the variance of the stationary solution of the model. In these cases, the Table 2.1 shows that the MDL procedure provides satisfactory results when there is one break in the model. But this procedure is not really efficient in the case of two breaks (see scenarios $\mathbf{A}_3$ and $\mathbf{A}_4$). In Table 2.3, we also consider two scenarios ($\mathbf{B}_2$ and $\mathbf{B}_3$) of AR(2) process where there is a change in the parameters but the variance of the stationary solution remains constant. As can be seen, the penalty $\widehat{\kappa}_n$ still works well whereas the MDL procedure provides poor results. Moreover, the Table 2.6 shows that, the MDL procedure provides sometimes excellent results (scenarios **A** and **J**), but also very weak result (scenario **C**).
Finally, one can see that if the $\widehat{\kappa}_n$ penalty does not always provide the best results, its results in all scenarios remain satisfactory, in the sense that in all considered scenarios, the estimated probability to fit the true number of break is greater than 0.50 for $n = 1000$. The use of our method with $\widehat{\kappa}_n$ is clearly the best possible trade-off for one break models. In the case of two breaks, the $\widehat{\kappa}_n$ penalty provides best results. Contrary to the MDL procedure, the QLIK criteria with $\widehat{\kappa}_n$ penalty works well in the AR models when the changes in the parameters does not induce a change in the variance of the stationary solution. For

Model		$\widehat{K}=1$ no break	$\widehat{K}=2$ one break	$\widehat{K}\geq 3$ more than 2 breaks
scenario A	$\kappa_n = \widehat{\kappa}_n$	0.560	0.390	0.050
$(K^*=1)$	$\kappa_n = \log n$	0.000	0.000	1.000
	$\kappa_n = \sqrt{n}$	0.600	0.390	0.010
	MDL procedure	0.958	0.042	0.000
scenario C	$\kappa_n = \widehat{\kappa}_n$	0.330	0.51	0.160
$(K^*=2)$	$\kappa_n = \log n$	0.000	0.000	1.000
	$\kappa_n = \sqrt{n}$	0.770	0.330	0.000
	MDL procedure	0.804	0.192	0.004
scenario J	$\kappa_n = \widehat{\kappa}_n$	0.050	0.630	0.320
$(K^*=2)$	$\kappa_n = \log n$	0.000	0.000	1.000
	$\kappa_n = \sqrt{n}$	0.280	0.620	0.100
	MDL procedure	0.008	0.952	0.040

Table 2.6 – Frequencies of the number of breaks estimated after 100 replications for GARCH$(1,1)$ processes with $n = 1000$ following the scenarios **A**, **C** and **J** of Davis *et al.* (2008) [25]. The results of MDL procedure were taken in Table I of [25].

all these reasons, we recommend to use our procedure with the penalty term $\kappa_n = \widehat{\kappa}_n$.

TARCH(1) models : we consider an example of problem (2.3) where X is a $TARCH(1)$ with one change :

$$X_t = \sigma_t\,\xi_t\ ,\ \sigma_t = b_0^*(j)+b_1^{+*}(j)\max(X_{t-1},0)-b_1^{-*}(j)\min(X_{t-1},0),\ \forall t\in T_j,\ j=1,2=K^*.$$

The vector of parameter is $\theta^* = (b_0^*, b_1^{+*}, b_1^{-*})$. Here we assume that the number of breaks is known, *i.e.* $K = K^* = 2$ but the break instant t^* and parameters θ^* are unknown. For $n = 1000$ and $n = 2000$, we generate 100 independent replications of $(X_1, \cdots, X_n)$ with parameters $\theta^*(1) = (0.01, 0.05, 0.04)$ for $t \leq t^* = 0.4n$ and $\theta^*(2) = (0.01, 0.05, 0.1)$ for $t > t^*$. Table 2.5 provides the sample mean and the standard deviation of $\widehat{\tau}_n$, the sample mean of the error $|\widehat{\tau}_n - \tau^*|$ and the sample means of $\widehat{\theta}_n(1)$ and $\widehat{\theta}_n(2)$.
We can see that the results obtained for AR(1) and GARCH$(1,1)$ models are much better than those obtained for TARCH(1) process even when K^* is known and $K^* = 2$ instead of $K^* = 3$. This is explained by the fact that this model provides an asymmetric function of the past observations. Thus, some asymmetric effects can be confused with breaks.
owever, Table 2.5 shows that the change is correctly detected and the decay rate of the error $|\widehat{\tau}_n - \tau^*|$ is confirmed. Figure 2.5 presents an example of such TARCH(1) process with one break.

2.5.4 Application to financial data : FTSE index analysis

Now we apply our detection of changes methodology to the series of the log-returns of the closing values of the FTSE index : the share index of the 100 most highly capitalized UK companies listed on the London Stock Exchange, with the aim of investigating whether and how any detected breakpoints correspond to the milestones of the recent financial crisis. This is a trajectory composed with $n = 1428$ observations ranging from 27 July 2005 to 18 March 2011, *i.e.* roughly 6 trading years and uploaded from Yahoo finance (see Figure 2.6). We studied the log-ratio of the closing daily prices. Remark that

	Mean of $\widehat{\tau}_n$ $\pm$ standard deviation	Mean of $\lvert\widehat{\tau}_n - \tau^*\rvert$	Mean of $\widehat{\theta}(1)$	Mean $\widehat{\theta}(2)$
$n = 1000$	0.436 ± 0.126	0.093	(0.056, 0.071, 0.044)	(0.067, 0.057, 0.103)
$n = 2000$	0.419 ± 0.063	0.044	(0.059, 0.052, 0.051)	(0.066, 0.061, 0.098)

TABLE 2.7 – The estimated parameters for a TARCH(1) process with one break from 100 independent replications. The parameter $\theta^*(1) = (0.01, 0.05, 0.04)$ changes to $\theta^*(2) = (0.01, 0.05, 0.1)$ at $t^* = 0.4n$.

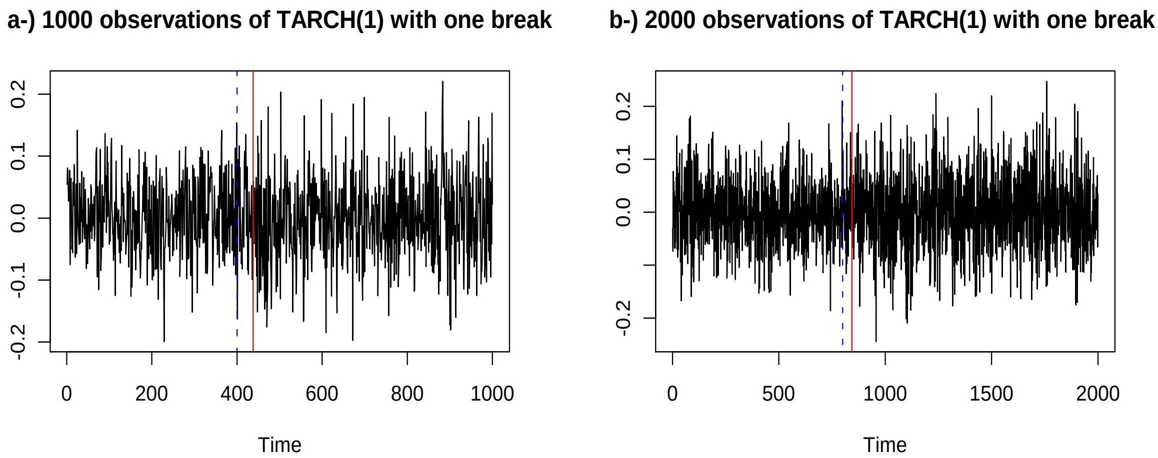

FIGURE 2.5 – Example of a the trajectory of TARCH(1) with one change (the red line represents the estimated break instant and the dotted line represents the true one).

we completely treat the period studied in [30].
The *penQLIK* contrast is applied for a GARCH$(1, 1)$ model (see (4.3) for a formal definition). The slope estimation procedure applied with $u_n = [n/(4 * \log(n))] = 49$ and $K_{max} = 25$ returns the values $\hat{\kappa} \approx 15$ and $\hat{K} = 4$, i.e. three breaks $\hat{t}_1 = 499$, $\hat{t}_2 = 792$ and $\hat{t}_3 = 853$. These values are close to the three breaks obtained in [30] :

$\hat{t}_1 = 499$, corresponding to 16 July 2007. From Wikipedia : "During the week of July 16, 2007, Bear Stearns disclosed that the two subprime hedge funds had lost nearly all of their value amid a rapid decline in the market for subprime mortgages."

$\hat{t}_2 = 792$, corresponding to 11 September 2008. From Wikipedia "On September 15, 2008, Lehman Brothers filed for Chapter 11 bankruptcy protection following the massive exodus of most of its clients, drastic losses in its stock, and devaluation of its assets by credit rating agencies".

$\hat{t}_3 = 853$, corresponding to the 5 December 2008. From Wikipedia "In the final quarter of 2008, the financial crisis saw the G-20 group of major economies assume a new significance as a focus of economic and financial crisis management."

Remark that our first two breaks are closer to the events identified in [30] than their own breaks. Analyzing the estimated values of coefficients (see Figure 2.7), breaks are due to changes of the coefficients a_1 and b_1 in the GARCH$(1, 1)$ model (4.3). There is no break for the mean μ and the a_0 coefficients, valued close to 0. Next, we compare the

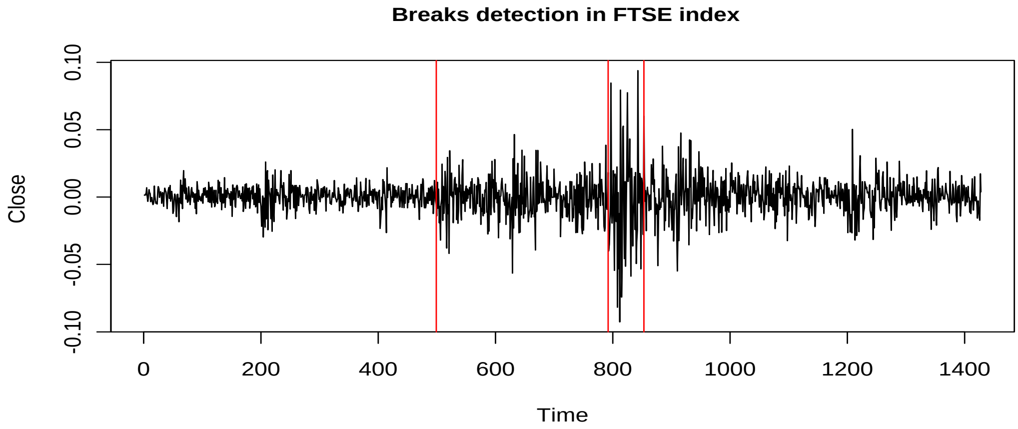

FIGURE 2.6 – The log-ratios of the closing daily price of the FTSE index. The vertical red lines represent the estimated instant of breaks.

fitted volatilities of the parameters estimations over the whole sequence and within the 4 periods. The third period, corresponding to a change of the value of the parameter a_1 ($a_1(3)$ is not significantly different from 0), leads to an estimated volatility satisfying the recurrence equation $\sigma_t \approx a_0(3) + b_1(3)\sigma_{t-1}$. In this period of high volatility, the estimated volatilities have different behaviors whether we take the break into account or not.

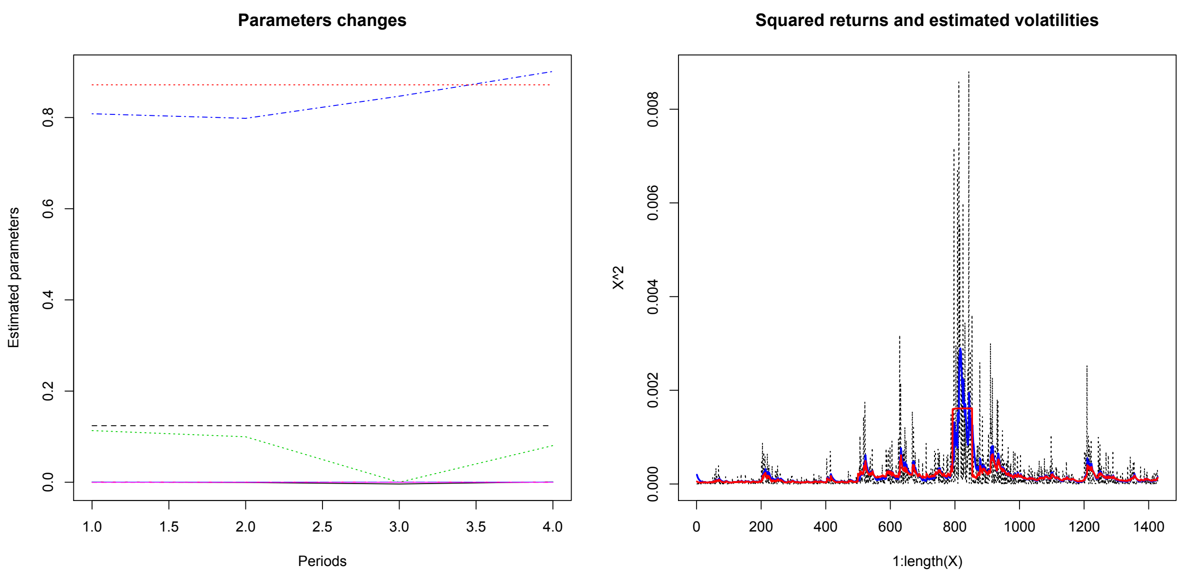

FIGURE 2.7 – The right plot represents the values of the parameters. The black line represents the values of μ, the red line of a_0, the green line of a_1 and the blue line of b_1. The left plot corresponds to the squared log-returns and the fitted volatilities, in blue with estimations over the whole sequence, in red with breaks.

2.6 Proofs of the main results

In the sequel C denotes a positive constant whom value may differ from one inequality to another and (v_n) is a sequence such that $v_n = n/\kappa_n$ for all $n \geq 1$.

2.6.1 Some preliminary result

The following technical lemma is useful in the sequel :

Lemma 2.6.1. Suppose that $\theta_j^* \in \Theta$ for $j = 1, \ldots, K^*$, Θ satisfying A with $r \geq 2$ and under the assumptions $A_0(f_\theta, \Theta)$, $A_0(M_\theta, \Theta)$ (or $A_0(h_\theta, \Theta)$) and $D(\Theta)$, then there exists $C > 0$ such that

$$\text{for all } t \in \mathbb{Z}, \quad \mathbb{E}\Big(\sup_{\theta \in \Theta} |q_t(\theta)|\Big) \leq C.$$

Proof Using the inequality $(a+b)^2 \leq 2(a^2+b^2)$, we have for all $t \in \mathbb{Z}$:

$$\begin{aligned}
\|f_\theta^t\|_\Theta^2 &\leq 2\Big(\|f_\theta^t - f_\theta(0,\ldots)\|_\Theta^2 + \|f_\theta(0,\ldots)\|_\Theta^2\Big) \\
&\leq 2\Big(\Big(\sum_{i\geq 1} \alpha_i^{(0)}(f_\theta, \Theta)\Big) \cdot \sum_{i \geq 1} \alpha_i^{(0)}(f_\theta, \Theta)|X_{t-i}|^2 + \|f_\theta(0,\ldots)\|_\Theta^2\Big),
\end{aligned}$$

therefore

$$\mathbb{E}\|f_\theta^t\|_\Theta^2 \leq 2\Big(C\Big(\sum_{i\geq 1} \alpha_i^{(0)}(f_\theta, \Theta)\Big)^2 + \|f_\theta(0,\ldots)\|_\Theta^2\Big).$$

Thus $\mathbb{E}\|f_\theta^t\|_\Theta^2 \leq C$ for all $t \in \mathbb{Z}$ and similarly $\mathbb{E}(\|h_\theta^t\|_\Theta) = \mathbb{E}(\|M_\theta^t\|_\Theta^2) \leq C$. Yet, under assumption $(D(\Theta))$, we have : $|q_t(\theta)| \leq \dfrac{1}{\underline{h}}|X_t - f_\theta^t|^2 + |\log(h_\theta^t)|$ and using inequality $\log x \leq x - 1$ for all $x > 0$, it follows :

$$|\log(h_\theta^t)| = \Big|\log(\underline{h}) + \log(\frac{h_\theta^t}{\underline{h}})\Big| \leq 1 + |\log(\underline{h})| + \frac{1}{\underline{h}} h_\theta^t.$$

Finally, we have for all $t \in \mathbb{Z}$:

$$\mathbb{E}\big(\sup_{\theta\in\Theta} |q_t(\theta)|\big) \leq 1 + |\log \underline{h}| + \frac{1}{\underline{h}}\big(\mathbb{E}\|h_\theta^t\|_\Theta + 2\mathbb{E}|X_t|^2 + 2\mathbb{E}\|f_\theta^t\|_\Theta^2\big) \leq C. \qquad \blacksquare$$

2.6.2 Comparison with stationary solutions

In the following, we assume that $\theta_j^* \in \Theta$ for all $j = 1, \ldots, K^*$, Θ satisfying A with $r \geq 1$. It comes from [8] that the equation

$$X_{t,j} = M_{\theta_j^*}\big((X_{t-k,j})_{k\in\mathbb{N}^*}\big) \cdot \xi_t + f_{\theta_j^*}\big((X_{t-k,j})_{k\in\mathbb{N}^*}\big) \quad \text{for all } t \in \mathbb{Z}$$

has r order stationary solution $\big(X_{t,j}\big)_{t\in\mathbb{Z}}$ for any $j = 1, \ldots, K^*$. Then

Lemma 2.6.2. Assume that the assumptions $A_0(f_\theta, \Theta)$, $A_0(M_\theta, \Theta)$ (or $A_0(h_\theta, \Theta)$) hold and that $\theta_j^* \in \Theta$ for $j = 1, \ldots, K^*$, Θ satisfying A for $r \geq 2$. Then :

1. $X_t = X_{t,1}$ for all $t \leq t_1^*$;
2. There exists $C > 0$ such that for any $j \in \{2, \cdots, K^*\}$, for all $t \in T_j^*$,

$$\begin{aligned}
\|X_t - X_{t,j}\|_r &\leq C\Big(\inf_{1\leq p \leq t - t_{j-1}^*} \Big\{\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p} + \sum_{i\geq p} \beta_i^{(0)}(\theta_j^*)\Big\}\Big) \\
\|X_t^2 - X_{t,j}^2\|_{r/2} &\leq C\Big(\inf_{1\leq p \leq t - t_{j-1}^*} \Big\{\tilde\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p} + \sum_{i\geq p} \tilde\beta_i^{(0)}(\theta_j^*)\Big\}\Big).
\end{aligned}$$

Proof. 1. It is obvious from the definition of X.
2. Let $j \in \{2, \cdots, K^*\}$, we proceed by induction on $t \in T_j^*$.
First consider the general case where $A_0(f_\theta, \{\theta\})$ and $A_0(M_\theta, \{\theta\})$ hold with $\beta^{(0)}(\theta) < 1$. By Proposition 2.2.2, there exists $C_r \geq 0$ such that $\|X_t - X_{t,j}\|_r \leq \|X_t\|_r + \|X_{t,j}\|_r \leq C + \max_{1 \leq j \leq K^*} \|X_{0,j}\|_r \leq C_r$ for all $j = 1, \ldots, K^*$ and $t \in \mathbb{Z}$. For $1 \leq p \leq t - t_{j-1}^*$ let $u_\ell := \sup_{t_{j-1}^* + \ell p \leq i \leq t_j^*} \|X_i - X_{i,j}\|_r$. Then $\|X_t - X_{t,j}\|_r \leq u_{[(t-t_{j-1}^*)/p]}$ and for any $t \leq i \leq t_j^*$:

$$\begin{aligned}
\|X_i - X_{i,j}\|_r &\leq \sum_{k \geq 1} \beta_k^{(0)}(\theta_j^*) \|X_{i-k} - X_{i-k,j}\|_r \\
&\leq \sum_{k=1}^{p} \beta_k^{(0)}(\theta_j^*) \|X_{i-k} - X_{i-k,j}\|_r + C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \\
&\leq \beta^{(0)}(\theta_j^*) u_{[(t-t_{j-1}^*)/p]-1} + C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*).
\end{aligned}$$

Similarly, it is easy to show that for all $1 \leq \ell \leq [(t - t_{j-1}^*)/p]$ we have

$$u_\ell \leq \beta^{(0)}(\theta_j^*) u_{\ell-1} + C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*).$$

Denote $a = \beta^{(0)}(\theta_j^*) < 1$, $b = C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*)$ such that $u_\ell \leq a u_{\ell-1} + b$. Considering $w_0 = u_0$ and $w_l = a w_{\ell-1} + b$, then $w_\ell = a^\ell w_0 + b(1 - a^{\ell-1})/(1-a) \leq a^\ell w_0 + b/(1-a)$. Since $u_0 \leq C_r$ by definition and $u_\ell \leq w_\ell$ for any ℓ, we have :

$$\begin{aligned}
u_\ell &\leq a^\ell u_0 + \frac{b}{1-a} \leq \big(\beta^{(0)}(\theta_j^*)\big)^\ell C_r + \frac{C_r}{1 - \beta^{(0)}(\theta_j^*)} \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \\
&\leq \frac{C_r}{1 - \beta^{(0)}(\theta_j^*)} \Big(\big(\beta^{(0)}(\theta_j^*)\big)^\ell + \sum_{k>p} \beta_k^{(0)}(\theta_j^*)\Big).
\end{aligned}$$

Thus for all $1 \leq p \leq t - t_{j-1}^*$

$$\|X_t - X_{t,j}\|_r \leq \beta^{(0)}(\theta_j^*) u_{[(t-t_{j-1}^*)/p]-1} + C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \leq C\big(\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p} + \sum_{i \geq p} \beta_i^{(0)}(\theta_j^*)\big)$$

and Lemma 2.6.2 is proved.

In the ARCH-type case when $f_\theta = 0$ and $A_0(h_\theta, \{\theta\})$ holds with $\tilde{\beta}^{(0)}(\theta) < 1$, we follow the same reasoning than previously starting from the inequality

$$\|X_i^2 - X_{i,j}^2\|_{r/2} \leq \sum_{k \geq 1} \tilde{\beta}_k^{(0)}(\theta_j^*) \|X_{i-k}^2 - X_{i-k,j}^2\|_{r/2}.$$

For all $j = 1, \ldots, K^*$ and $t \in \mathbb{Z}$, by Proposition 2.2.2, $\|X_i^2 - X_{i,j}^2\|_{r/2} \leq C_r^2$ and therefore

$$\tilde{u}_\ell \leq \tilde{\beta}^{(0)}(\theta_j^*) \tilde{u}_{\ell-1} + C_r^2 \sum_{k>p} \tilde{\beta}_k^{(0)}(\theta_j^*)$$

for $\tilde{u}_\ell = \sup_{t_{j-1}^* + \ell p \leq i \leq t_j^*} \|X_i^2 - X_{i,j}^2\|_{r/2}$ and Lemma 2.6.2 is proved. □

2.6.3 The asymptotic behavior of the likelihood

For the process $(X_{t,j})_{t\in T_j^*,\, j=1,\dots,K^*}$, for any $j\in\{1,\cdots,K^*\}$ and $s\in T_j^*$ denote :

$$q_{s,j}(\theta):=\frac{\big(X_{s,j}-f_\theta^{s,j}\big)^2}{h_\theta^{s,j}}+\log\big(h_\theta^{s,j}\big) \qquad (2.11)$$

with $f_\theta^{s,j}:=f_\theta(X_{s-1,j},X_{s-2,j},\dots)$, $M_\theta^{s,j}:=M_\theta(X_{s-1,j},X_{s-2,j},\dots)$ and $h_\theta^{s,j}:=(M_\theta^{s,j})^2$. For any $T\subset T_j^*$, denote

$$L_{n,j}(T,\theta):=-\frac{1}{2}\sum_{s\in T}q_{s,j}(\theta)$$

the likelihood of the j^{th} stationary model computed on T.

Lemma 2.6.3. Assume that $D(\Theta)$ holds.

1. If the assumption H_0 with $r\geq 2$ holds then for all $j=1,\dots,K^*$:

$$\frac{v_{n_j^*}}{n_j^*}\Big\|L_n(T_j^*,\theta)-L_{n,j}(T_j^*,\theta)\Big\|_\Theta\xrightarrow[n\to\infty]{\texttt{a.s.}}0.$$

2. For $i=1,2$, if the assumption H_i with $r\geq 4$ holds then for all $j=1,\dots,K^*$:

$$\frac{v_{n_j^*}}{n_j^*}\Big\|\frac{\partial^i L_n(T_j^*,\theta)}{\partial\theta^i}-\frac{\partial^i L_{n,j}(T_j^*,\theta)}{\partial\theta^i}\Big\|_\Theta\xrightarrow[n\to\infty]{\texttt{a.s.}}0.$$

Proof. 1-) For any $\theta\in\Theta$, $\Big|\frac{1}{n_j^*}L_n(T_j^*,\theta)-\frac{1}{n_j^*}L_{n,j}(T_j^*,\theta)\Big|\leq\dfrac{1}{n_j^*}\sum\limits_{k=1}^{n_j^*}|q_{t_{j-1}^*+k}(\theta)-q_{t_{j-1}^*+k,j}(\theta)|$. Then :

$$v_{n_j^*}\Big\|\frac{1}{n_j^*}L_n(T_j^*,\theta)-\frac{1}{n_j^*}L_{n,j}(T_j^*,\theta)\Big\|_\Theta\leq\frac{v_{n_j^*}}{n_j^*}\sum_{k=1}^{n_j^*}\|q_{t_{j-1}^*+k}(\theta)-q_{t_{j-1}^*+k,j}(\theta)\|_\Theta.$$

By Corollary 1 of Kounias and Weng [47], with $r\leq 4$ and no loss of generality, the proof of Lemma 2.6.3 1-) is achieved if

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}I\!\!E\big(\|q_{t_{j-1}^*+k}(\theta)-q_{t_{j-1}^*+k,j}(\theta)\|_\Theta^{r/4}\big)<\infty. \qquad (2.12)$$

Let us prove (2.12). For any $\theta\in\Theta$, we have :

$$\begin{aligned}|q_s(\theta)-q_{s,j}(\theta)|&\leq\frac{1}{\underline{h}^2}|X_s-f_\theta^s|^2|h_\theta^s-h_\theta^{s,j}|\\&+\frac{1}{\underline{h}}\big(|X_s^2-X_{s,j}^2|+|f_\theta^s-f_\theta^{s,j}||f_\theta^s+f_\theta^{s,j}+2X_s|+2|f_\theta^{s,j}||X_s-X_{s,j}|+|h_\theta^s-h_\theta^{s,j}|\big). \qquad (2.13)\end{aligned}$$

First consider the general case with $A_0(f_\theta,\{\theta\})$ and $A_0(M_\theta,\{\theta\})$ hold and $\beta^{(0)}(\theta)<1$:

$$\begin{aligned}\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta\leq C\big(1&+|X_{s,j}|+|X_s|^2+\|f_\theta^{s,j}\|_\Theta+\|f_\theta^s\|_\Theta^2\big)\\&\times\big(|X_s-X_{s,j}|+\|f_\theta^s-f_\theta^{s,j}\|_\Theta+\|h_\theta^s-h_\theta^{s,j}\|_\Theta\big),\end{aligned}$$

and by Cauchy-Schwartz Inequality,

$$(\mathbb{E}\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4})^2 \le C\mathbb{E}\big[\big(1+|X_{s,j}|+|X_s|^2+\|f_\theta^{s,j}\|_\Theta+\|f_\theta^s\|_\Theta^2\big)^{r/2}\big] \\ \times \mathbb{E}\big[\big(|X_s-X_{s,j}|+\|f_\theta^s-f_\theta^{s,j}\|_\Theta+\|h_\theta^s-h_\theta^{s,j}\|_\Theta\big)^{r/2}\big].$$

Using Proposition (2.2.2) and the argument of the proof of Lemma (2.6.1) we claim that $\mathbb{E}|X_s|^r \le C$, $\mathbb{E}\|f_\theta^s\|_\Theta^r \le C$ and that $\mathbb{E}\|f_\theta^{s,j}\|_\Theta^r \le C$. Thus :

$$(\mathbb{E}\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4})^2 \le C\big(\mathbb{E}|X_s-X_{s,j}|^{r/2}+\mathbb{E}\|f_\theta^s-f_\theta^{s,j}\|_\Theta^{r/2}+\mathbb{E}\|h_\theta^s-h_\theta^{s,j}\|_\Theta^{r/2}\big). \quad (2.14)$$

Since $r/2 \ge 1$, we will use the $L^{r/2}$ norm. By Lemma 2.6.2 :

$$\|X_s-X_{s,j}\|_{r/2} \le \|X_s-X_{s,j}\|_r \le C \inf_{1\le p\le k}\{\beta^{(0)}(\theta_j^*)^{k/p}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\} \\ \le C \inf_{1\le p\le k/2}\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\}.$$

$$\implies \quad \mathbb{E}|X_s-X_{s,j}|^{r/2} \le C\Big(\inf_{1\le p\le k}\{\beta^{(0)}(\theta_j^*)^{k/p}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\}\Big)^{r/2}. \quad (2.15)$$

Moreover, as $(A_0(M_\theta,\Theta))$ holds, we have :

$$\|\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|_{r/2} \le C\sum_{i\ge 1}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r. \quad (2.16)$$

From (2.16) we obtain :

$$\|\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|_{r/2} \le C\Big(\sum_{i=1}^{k/2-1}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r+\sum_{i\ge k/2}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r\Big).$$

For all $s \ge t_{j-1}^*$ and $1 \le i \le k/2-1$, then $s-i > t_{j-1}^*$, $s-i > k/2$ and by Lemma 2.6.2 :

$$\begin{aligned} \|X_{s-i}-X_{s-i,j}\|_r &\le C \inf_{1\le p\le k-i}\{\beta^{(0)}(\theta_j^*)^{(k-i)/p}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\} \\ &\le C \inf_{1\le p\le k/2}\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\} \end{aligned}$$

Thus, we can find $C>0$ not depending on s such that :

$$\mathbb{E}\|h_\theta^s-h_\theta^{s,j}\|_\Theta^{r/2} \le C\Big(\inf_{1\le p\le k/2}\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\}+\sum_{i\ge k/2}\alpha_i^{(0)}(M_\theta,\Theta)\Big)^{r/2}. \quad (2.17)$$

Similarly, we obtain :

$$\mathbb{E}\|f_\theta^s-f_\theta^{s,j}\|_\Theta^{r/2} \le C\Big(\inf_{1\le p\le k/2}\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\}+\sum_{i\ge k/2}\alpha_i^{(0)}(f_\theta,\Theta)\Big)^{r/2}. \quad (2.18)$$

Relations (2.14), (2.15), (2.17) et (2.18) give (the same inequality holds with h_θ replaced by M_θ) :

$$\mathbb{E}\, \|q_s(\theta) - q_{s,j}(\theta)\|_\Theta^{r/4} \leq C\Big[\Big(\inf_{1\leq p\leq k/2}\{\beta^{(0)}(\theta_j^*)^{k/(2p)} + \sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\}\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha_i^{(0)}(f_\theta,\Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha_i^{(0)}(M_\theta,\Theta)\Big)^{r/4}\Big]. \quad (2.19)$$

By definition $u_k = kc^*/\log(k)$ ($\leq k/2$ for large value of k) satisfies the relation

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\big(\beta^{(0)}(\theta_j^*)\big)^{rk/8u_k} < \infty.$$

Choosing $p = u_k$ in (2.19) we obtain :

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\mathbb{E}(\|q_{t_{j-1}^*+k}(\theta) - q_{t_{j-1}^*+k,j}(\theta)\|_\Theta^{r/4}) \leq \sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\big(\beta^{(0)}(\theta_j^*)\big)^{rk/8u_k}$$
$$+\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\geq u_k}\beta_i^{(0)}(\theta_j^*)\Big)^{r/4} + \sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\geq k/2}(\alpha_i^{(0)}(f_\theta,\Theta)+\alpha_i^{(0)}(M_\theta,\Theta))\Big)^{r/4}.$$

This bound is finite by assumption and therefore (2.12) is established.

In the ARCH-type case when $f_\theta = 0$ and $A_0(h_\theta,\{\theta\})$ holds with $\tilde\beta^{(0)}(\theta) < 1$, we follow the same reasoning than previously remarking that (2.13) has the simplified form :

$$|q_s(\theta) - q_{s,j}(\theta)| \leq \frac{1}{\underline{h}^2}X_s^2|h_\theta^s - h_\theta^{s,j}| + \frac{1}{\underline{h}}|X_s^2 - X_{s,j}^2| + \frac{1}{\underline{h}}|h_\theta^s - h_\theta^{s,j}|.$$

Then

$$(\mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|_\Theta^{r/4})^2 \leq C\mathbb{E}\big[(|X_s^2 - X_{s,j}^2| + \|h_\theta^s - h_\theta^{s,j}\|_\Theta)^{r/2}\big].$$

As $\|\|h_\theta^s - h_\theta^{s,j}\|_\Theta\|_{r/2} \leq C\sum_{i\geq 1}\alpha_i^{(0)}(h_\theta,\Theta)\|X_{s-i}^2 - X_{s-i,j}^2\|_{r/2}$ we derive from Lemma 2.6.2,

$$\mathbb{E}\, \|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4} \leq C\Big[\Big(\inf_{1\leq p\leq k/2}\{\tilde\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\geq p}\tilde\beta_i^{(0)}(\theta_j^*)\}\Big)^{r/4}+\Big(\sum_{i\geq k/2}\alpha_i^{(0)}(h_\theta,\Theta)\Big)^{r/4}\Big].$$

We easily conclude to the result by choosing $p = u_k$ as above.

2-) We detail the proof for one order derivation in the general case where $A_0(f_\theta,\{\theta\})$ and $A_0(M_\theta,\{\theta\})$ hold with $\beta^{(0)}(\theta) < 1$. The proofs of the other cases follow the same reasoning.
Let $j \in \{1,\cdots,K^*\}$ and $i = 1,\cdots,d$, we have :

$$\frac{v_{n_j^*}}{n_j^*}\Big\|\frac{\partial L_n(T_j^*,\theta)}{\partial\theta_i} - \frac{\partial L_{n,j}(T_j^*,\theta)}{\partial\theta_i}\Big\|_\Theta \leq \frac{v_{n_j^*}}{n_j^*}\sum_{k=1}^{n_j^*}\Big\|\frac{\partial q_{t_{j-1}^*+k}(\theta)}{\partial\theta_i} - \frac{\partial q_{t_{j-1}^*+k,j}(\theta)}{\partial\theta_i}\Big\|_\Theta.$$

As previously, using Corollary 1 of [47], when $r \leq 4$ with no loss of generality, Lemma 2.6.3 2-) will be established if

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\mathbb{E}\Big(\Big\|\frac{\partial q_{t_{j-1}^*+k}(\theta)}{\partial\theta_i} - \frac{\partial q_{t_{j-1}^*+k,j}(\theta)}{\partial\theta_i}\Big\|_\Theta^{r/4}\Big) < \infty. \quad (2.20)$$

For any $s \geq t^*_{j-1}$ denote $k = s - t^*_{j-1}$. For any $\theta \in \Theta$, we have :

$$\frac{\partial q_s(\theta)}{\partial \theta_i} = -2\frac{(X_s - f^s_\theta)}{h^s_\theta}\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{(X_s - f^s_\theta)^2}{(h^s_\theta)^2}\frac{\partial h^s_\theta}{\partial \theta_i} + \frac{1}{h^s_\theta}\frac{\partial h^s_\theta}{\partial \theta_i}$$

$$\frac{\partial q_{s,j}(\theta)}{\partial \theta_i} = -2\frac{(X_{s,j} - f^{s,j}_\theta)}{h^{s,j}_\theta}\frac{\partial f^{s,j}_\theta}{\partial \theta_i} - \frac{(X_{s,j} - f^{s,j}_\theta)^2}{(h^{s,j}_\theta)^2}\frac{\partial h^{s,j}_\theta}{\partial \theta_i} + \frac{1}{h^{s,j}_\theta}\frac{\partial h^{s,j}_\theta}{\partial \theta_i}.$$

Thus, using $|a_1b_1c_1 - a_2b_2c_2| \leq |a_1 - a_2||b_2||c_2| + |b_1 - b_2||a_1||c_2| + |c_1 - c_2||a_1||b_1|$,

$$\begin{aligned}\Big\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\Big\|_\Theta &\leq 2\Big(\frac{1}{\underline{h}^2}\|h^s_\theta - h^{s,j}_\theta\|_\Theta\|X_{s,j} - f^{s,j}_\theta\|_\Theta\Big\|\frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta \\
&+\frac{1}{\underline{h}}(|X_s - X_{s,j}| + \|f^s_\theta - f^{s,j}_\theta\|_\Theta)\Big\|\frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta + \frac{1}{\underline{h}}\Big\|\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta\|X_s - f^s_\theta\|_\Theta\Big) \\
&+\frac{2}{\underline{h}^3}\|h^s_\theta - h^{s,j}_\theta\|_\Theta\|X_{s,j} - f^{s,j}_\theta\|^2_\Theta\Big\|\frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta \\
&+\frac{1}{\underline{h}}(|X_s - X_{s,j}| + \|f^s_\theta - f^{s,j}_\theta\|_\Theta)(|X_s + X_{s,j}| + \|f^s_\theta + f^{s,j}_\theta\|_\Theta)\Big\|\frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta \\
&+\frac{1}{\underline{h}^2}\Big\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta\|X_s - f^s_\theta\|^2_\Theta + \frac{1}{\underline{h}^2}\|h^s_\theta - h^{s,j}_\theta\|_\Theta\Big\|\frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta + \frac{1}{\underline{h}}\Big\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta\end{aligned}$$

So for all $s \geq t^*_{j-1}$ it holds :

$$\begin{aligned}&\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\|_\Theta \\
&\leq C\Big(1 + |X_s|^2 + |X_{s,j}|^2 + \|f^s_\theta\|^2_\Theta + \|f^{s,j}_\theta\|^2_\Theta + \|\frac{\partial f^s_\theta}{\partial \theta_i}\|^2_\Theta + \|\frac{\partial f^{s,j}_\theta}{\partial \theta_i}\|^2_\Theta + \|\frac{\partial h^s_\theta}{\partial \theta_i}\|^2_\Theta + \|\frac{\partial h^{s,j}_\theta}{\partial \theta_i}\|^2_\Theta\Big) \\
&\times\Big(|X_s - X_{s,j}| + \|f^s_\theta - f^{s,j}_\theta\|_\Theta + \|h^s_\theta - h^{s,j}_\theta\|_\Theta + \|\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{\partial f^{s,j}_\theta}{\partial \theta_i}\|_\Theta + \Big\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta\Big)\end{aligned}$$

Since the processes admits finite moments of order r, by Cauchy-Schwartz Inequality :

$$\begin{aligned}\Big(\mathbb{E}\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\|^{r/4}_\Theta\Big)^2 &\leq C\Big(\mathbb{E}|X_s - X_{s,j}|^{r/2} + \mathbb{E}(\|f^s_\theta - f^{s,j}_\theta\|^{r/2}_\Theta) + \mathbb{E}(\|h^s_\theta - h^{s,j}_\theta\|^{r/2}_\Theta) \\
&+ \mathbb{E}\|\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{\partial f^{s,j}_\theta}{\partial \theta_i}\|^{r/2}_\Theta + \mathbb{E}\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\|^{r/2}_\Theta\Big)\end{aligned}$$

As $(A_0(M_\theta, \Theta))$ and $(A_1(M_\theta, \Theta))$ hold necessarily in this case, with the arguments of the proof of 1-), for all $s \geq t^*_{j-1}$,

$$\begin{aligned}\mathbb{E}\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\|^{r/4}_\Theta &\leq C\Big[\Big(\inf_{1\leq p\leq k/2}\{\beta^{(0)}(\theta^*_j)^{k/(2p)} + \sum_{i\geq p}\beta^{(0)}_i(\theta^*_j)\}\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(f_\theta, \Theta)\Big)^{r/4} \\
&+ \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(M_\theta, \Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(1)}_i(f_\theta, \Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(1)}_i(M_\theta, \Theta)\Big)^{r/4}\Big]\end{aligned}$$

Choosing $p = u_k = kc^*/\log(k)$, we show (as in proof of 1-)) that :

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\mathbb{E}\Big(\Big\|\frac{\partial q_{t^*_{j-1}+k}(\theta)}{\partial \theta_i} - \frac{\partial q_{t^*_{j-1}+k,j}(\theta)}{\partial \theta_i}\Big\|^{r/4}_\Theta\Big) < \infty.$$

Therefore (2.20) is proved and Lemma 2.6.3 2-) also. $\square$

2.6.4 Consistency when the breaks are known

When the breaks are known, we can choose $v_n = 1$ for all n ; in (2.4), the penalty term does not matter at all.

Proposition 2.6.1. For all $j = 1, \dots, K^*$, under the assumptions of Lemma 2.6.3 1-) with $v_n = 1$ for all n, if the assumption $\mathrm{Id}(\Theta)$ holds then

$$\widehat{\theta}_n(T_j^*) \xrightarrow[n\to\infty]{\text{a.s.}} \theta_j^*.$$

Proof. Let us first give the following useful corollary of Lemma 2.6.3

Corollary 2.6.1. i-) under the assumptions of Lemma 2.6.3 1-) we have :

$$\Big\| \frac{1}{n_j^*} \widehat{L}_n(T_j^*, \theta) - \mathcal{L}_j(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0 \ \text{ with } \ \mathcal{L}_j(\theta) = -\frac{1}{2} \mathbb{E}\left(q_{0,j}(\theta)\right).$$

ii-) Under assumptions of Lemma 2.6.3 2-) we have :

$$\Big\| \frac{1}{n_j^*} \frac{\partial^i \widehat{L}_n(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} \Big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0 \ \text{ with } \ \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} = -\frac{1}{2} \mathbb{E}\left(\frac{\partial^i q_{0,j}(\theta)}{\partial \theta^i}\right).$$

We conclude the proof of Proposition 2.6.1 using $\mathcal{L}_j(\theta) = -\frac{1}{2} \mathbb{E}\left(q_{0,j}(\theta)\right)$ has a unique maximum in θ_j^* (see [38]). From the almost sure convergence of the quasi-likelihood in i-) of Corollary 2.6.1, it comes :

$$\widehat{\theta}_n(T_j^*) = \underset{\theta \in \Theta}{\mathrm{Argmax}} \left(\frac{1}{n_j^*} \widehat{L}_n(T_j^*, \theta) \right) \xrightarrow[n\to\infty]{\text{a.s.}} \theta_j^*. \ \square$$

Proof of Corollary 2.6.1 Note that the proof of Lemma 2.6.3 can be repeated by replacing L_n by the quasi-likelihood $\widehat{L}_n$. Thus, we obtain for $i = 0, 1, 2$,

$$\frac{v_{n_j^*}}{n_j^*} \left\| \frac{\partial^i \widehat{L}_n(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i L_{n,j}(T_j^*, \theta)}{\partial \theta^i} \right\|_\Theta \xrightarrow[n\to\infty]{} 0. \tag{2.21}$$

i-) Let $j \in 1, \cdots, K^*$. From [8], we have :

$$\Big\| \frac{1}{n_j^*} L_{n,j}(T_j^*, \theta) - \mathcal{L}_j(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Using (2.21), the convergence to the limit likelihood follows.

ii-) From Lemma 4 and Theorem 1 of [8], $\Big\| \frac{1}{n_j^*} \frac{\partial^i L_{n,j}(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} \Big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0$ for $i = 1,\ 2$ and we conclude from (2.21). $\square$

2.6.5 Proof of Theorem 2.3.1

This proof is divided into two parts. In **part (1)** K^* is assumed to be known and we show $(\widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (\underline{\tau}^*, \underline{\theta}^*)$. In **part (2)**, K^* is unknown and we show $\widehat{K}_n \xrightarrow[n\to\infty]{\mathcal{P}} K^*$ which ends the proof of Theorem 2.3.1.

Part (1). Assume that K^* is known and denote for any $\underline{t} \in \mathcal{F}_{K^*}$:

$$\widehat{I}_n(\underline{t}) := \widehat{J}_n(K^*, \underline{t}, \widehat{\underline{\theta}}_n(\underline{t})) = -2 \sum_{k=1}^{K^*} \sum_{j=1}^{K^*} \widehat{L}_n \Big(T_k \cap T_j^*, \widehat{\theta}_n(T_k) \Big)$$

It comes that $\widehat{\underline{t}}_n = \underset{\underline{t} \in \mathcal{F}_{K^*}}{\text{Argmin}} \Big(\widehat{I}_n(\underline{t}) \Big)$. We show that $\widehat{\underline{\tau}}_n \xrightarrow[n\to\infty]{\mathrm{P}} \underline{\tau}^*$ as it implies $\widehat{\theta}_n(\widehat{T}_{n,j}) - \widehat{\theta}_n(T_j^*) \xrightarrow[n\to\infty]{\mathrm{P}} 0$ and from Proposition 2.6.1 $\widehat{\theta}_n(\widehat{T}_{n,j}) \xrightarrow[n\to\infty]{\mathrm{P}} \theta_j^*$ for all $j = 1, \cdots, K^*$. Without loss of generality, assume that $K^* = 2$ and let (u_n) be a sequence of positive integers satisfying $u_n \to \infty$, $u_n/n \to 0$ and for some $0 < \eta < 1$

$$\begin{aligned} V_{\eta,u_n} &= \{ \, t \in \mathbb{Z}/ \ \ |t - t^*| > \eta n \ ; \ u_n \leq t \leq n - u_n \, \}, \\ W_{\eta,u_n} &= \{ \, t \in \mathbb{Z}/ \ \ |t - t^*| > \eta n \ ; \ 0 < t < u_n \ \text{ or } \ n - u_n < t \leq n \, \}. \end{aligned}$$

Asymptotically, we have $\mathbb{P}(\|\widehat{\underline{\tau}}_n - \underline{\tau}^*\|_m > \eta) \simeq \mathbb{P}(|\widehat{t}_n - t^*| > \eta n)$. But

$$\begin{aligned} \mathbb{P}(|\widehat{t}_n - t^*| > \eta n) &\leq \mathbb{P}\Big(\widehat{t}_n \in V_{\eta,u_n} \Big) + \mathbb{P}\Big(\widehat{t}_n \in W_{\eta,u_n} \Big) \\ &\leq \mathbb{P}\Big(\min_{t \in V_{\eta,u_n}} (\widehat{I}_n(t) - \widehat{I}_n(t^*)) \leq 0 \Big) + \mathbb{P}\Big(\min_{t \in W_{\eta,u_n}} (\widehat{I}_n(t) - \widehat{I}_n(t^*)) \leq 0 \Big) \end{aligned}$$

we show with similar arguments that these two probabilities tend to 0. We only detail below the proof of $\mathbb{P}\Big(\min_{t \in V_{\eta,u_n}} (\widehat{I}_n(t) - \widehat{I}_n(t^*)) \leq 0 \Big) \to 0$ for shortness.

Let $t \in V_{\eta,u_n}$ satisfying $t^* \leq t$ (with no loss of generality), then $T_1 \cap T_1^* = T_1^*$, $T_2 \cap T_1^* = \emptyset$ and $T_2 \cap T_2^* = T_2$. We decompose :

$$\begin{aligned} \widehat{I}_n(t) - \widehat{I}_n(t^*) = 2 \Big(& \widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*)) - \widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)) \\ & - \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2)) \Big). \end{aligned} \tag{2.22}$$

As $\# T_1^* = t^*$, $\#(T_1 \cap T_2^*) = t - t^*$, $\# T_2 = n - t \geq u_n$, each term tends to ∞ with n. Using Proposition 2.6.1 and Corollary 2.6.1, we get the following convergence, uniformly on V_{η,u_n},

$$\widehat{\theta}_n(T_1^*) \xrightarrow[n\to\infty]{a.s.} \theta_1^*, \ \widehat{\theta}_n(T_2^*) \xrightarrow[n\to\infty]{a.s.} \theta_2^*, \ \widehat{\theta}_n(T_2) \xrightarrow[n\to\infty]{a.s.} \theta_2^* \ \text{ and } \ \Big\| \frac{\widehat{L}_n(T_1^*, \theta)}{n} - \tau_1^* \mathcal{L}_1(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0,$$

$$\Big\| \frac{\widehat{L}_n(T_1 \cap T_2^*, \theta)}{t - t^*} - \mathcal{L}_2(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0, \ \ \Big\| \frac{\widehat{L}_n(T_2, \theta)}{n - t} - \mathcal{L}_2(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

For any $\varepsilon > 0$, there exists an integer N_0 such that for any $n > N_0$,

$$\Big\| \frac{\widehat{L}_n(T_1^*, \theta)}{n} - \tau_1^* \mathcal{L}_1(\theta) \Big\|_\Theta < \frac{\varepsilon}{6}; \ \Big\| \frac{\widehat{L}_n(T_1 \cap T_2^*, \theta)}{t - t^*} - \mathcal{L}_2(\theta) \Big\|_\Theta < \frac{\varepsilon}{6}; \ \Big| \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*))}{n} - \tau_1^* \mathcal{L}_1(\theta_1^*) \Big| < \frac{\varepsilon}{6}$$

$$\Big| \frac{\widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*))}{t - t^*} - \mathcal{L}_2(\theta_2^*) \Big| < \frac{\varepsilon}{6}; \quad \frac{n - t}{n} \Big| \frac{\widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2))}{n - t} \Big| < \frac{\varepsilon}{6}$$

Thus, for $n > N_0$,

$$\begin{aligned} \tau_1^* \mathcal{L}_1(\theta_1^*) - \tau_1^* \mathcal{L}_1(\widehat{\theta}_n(T_1)) &= \tau_1^* \mathcal{L}_1(\theta_1^*) - \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*))}{n} + \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*))}{n} - \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1))}{n} \\ &\quad + \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1))}{n} - \tau_1^* \mathcal{L}_1(\widehat{\theta}_n(T_1)) \\ &\leq \frac{\varepsilon}{6} + \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*))}{n} - \frac{\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1))}{n} + \frac{\varepsilon}{6}. \end{aligned}$$

Then,

$$\frac{\widehat{L}_n(T_1^*,\widehat{\theta}_n(T_1^*))}{n}-\frac{\widehat{L}_n(T_1^*,\widehat{\theta}_n(T_1))}{n} \quad > \quad \tau_1^*\Big(\mathcal{L}_1(\theta_1^*)-\mathcal{L}_1(\widehat{\theta}_n(T_1))\Big)-\frac{\varepsilon}{3}. \qquad (2.23)$$

Similarly, for $n>N_0$:

$$\frac{\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{n}-\frac{\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_1))}{n}>\eta\Big(\mathcal{L}_2(\theta_2^*)-\mathcal{L}_2(\widehat{\theta}_n(T_1))\Big)-\frac{\varepsilon}{3}. \quad (2.24)$$

Finally, for $n>N_0$,

$$\frac{\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_2,\widehat{\theta}_n(T_2))}{n} \quad > \quad -\frac{\varepsilon}{6}, \qquad (2.25)$$

and from (2.22) and inequalities (2.23), (2.24) and (2.25) we obtain uniformly in t :

$$\frac{\widehat{I}_n(t)-\widehat{I}_n(t^*)}{n}>\tau_1^*\Big(\mathcal{L}_1(\theta_1^*)-\mathcal{L}_1(\widehat{\theta}_n(T_1))\Big)+\eta\Big(\mathcal{L}_2(\theta_2^*)-\mathcal{L}_2(\widehat{\theta}_n(T_1))\Big)-\frac{5}{6}\varepsilon, \quad n>N_0.$$

Since $\theta_1^*\neq\theta_2^*$, let $\mathcal{V}_1$, $\mathcal{V}_2$ be two open neighborhoods and disjoint of θ_1^* and θ_2^* respectively,

$$\delta_i:=\operatorname*{Inf}_{\theta\in\mathcal{V}_i^c}\Big(\mathcal{L}_i(\theta_i^*)-\mathcal{L}_i(\theta)\Big)>0 \quad \text{for} \quad i=1,2,$$

since the function $\theta\mapsto\mathcal{L}_j(\theta)$ has a strict maximum in θ_j^* (see [38]). With $\varepsilon=\min(\tau_1^*\delta_1,\ \eta\delta_2)$, we get

- if $\widehat{\theta}_n(T_1)\in\mathcal{V}_1$ i.e. $\widehat{\theta}_n(T_1)\in\mathcal{V}_2^c$, then $\dfrac{\widehat{I}_n(t)-\widehat{I}_n(t^*)}{n}>\eta\delta_2-\dfrac{5}{6}\varepsilon\geq\dfrac{\varepsilon}{6}$;
- If $\widehat{\theta}_n(T_1)\notin\mathcal{V}_1$ i.e. $\widehat{\theta}_n(T_1)\in\mathcal{V}_1^c$, then $\dfrac{\widehat{I}_n(t)-\widehat{I}_n(t^*)}{n}>\tau_1^*\delta_1-\dfrac{5}{6}\varepsilon\geq\dfrac{\varepsilon}{6}$.

In any case we prove that $\widehat{I}_n(t)-\widehat{I}_n(t^*)>\dfrac{\varepsilon}{6}n$ for $n>N_0$ and all $t\in V_{\eta,u_n}$. It implies that $\mathbb{P}\Big(\min\limits_{\underline{t}\in V_{\eta,u_n}}(\widehat{I}_n(t)-\widehat{I}_n(t^*))\leq 0\Big)\underset{n\to\infty}{\longrightarrow}0$ and we show similarly $\mathbb{P}\Big(\min\limits_{\underline{t}\in W_{\eta,u_n}}(\widehat{I}_n(t)-\widehat{I}_n(t^*))\leq 0\Big)\underset{n\to\infty}{\longrightarrow}0$. It follows directly that $\mathbb{P}(\|\underline{\widehat{\tau}}_n-\underline{\tau}^*\|_m>\eta)\underset{n\to\infty}{\longrightarrow}0$ for all $\eta>0$.

Part(2). Now K^* is unknown. For $K\geq 2$, $x=(x_1,\cdots,x_{K-1})\in\mathbb{R}^{K-1}$, $y=(y_1,\cdots,y_{K^*-1})\in\mathbb{R}^{K^*-1}$, denote

$$\|x-y\|_\infty=\max_{1\leq j\leq K^*-1}\ \min_{1\leq k\leq K-1}|x_k-y_j|.$$

The following Lemma follows directly from **Part(1)** and the definition of $\|\cdot\|_\infty$:

Lemma 2.6.4. Let $K\geq 1$, $(\underline{\widehat{t}}_n,\underline{\widehat{\theta}}_n)$ obtained by the minimization of $\widehat{J}_n(\underline{t},\underline{\theta})$ on $\mathcal{F}_K\times\Theta^K$ and $\underline{\widehat{\tau}}_n=\underline{\widehat{t}}_n/n$. Under assumptions of Theorem 2.3.1, $\|\underline{\widehat{\tau}}_n-\underline{\tau}^*\|_\infty\overset{\mathrm{P}}{\underset{n\to+\infty}{\longrightarrow}}0$ if $K\geq K^*$.

Now we use the following Lemma 2.6.5 which is proved below (see also [51]) :

Lemma 2.6.5. Under the assumptions of Lemma 2.6.3 i-), for any $K\geq 2$, there exists $C_K>0$ such that :

$$\forall(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta^K, \quad e_n(\underline{t},\underline{\theta})=2\sum_{j=1}^{K^*}\sum_{k=1}^{K}\frac{n_{kj}}{n}(\mathcal{L}_j(\theta_j^*)-\mathcal{L}_j(\theta_k))\geq\frac{C_K}{n}\|\underline{t}-\underline{t}^*\|_\infty.$$

Continue with the proof of **Part(2)** shared in two parts, *i.e.* we show that $P(\widehat{K}_n = K) \underset{n\to+\infty}{\longrightarrow} 0$ for $K < K^*$ and $K^* < K \leq K_{\max}$ separately. In any case, we have

$$\begin{aligned} P(\widehat{K}_n = K) &\leq P\Big(\inf_{(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta^K}(\widetilde{J}_n(K,\underline{t},\underline{\theta})) \leq \widetilde{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)\Big) \\ &\leq P\Big(\inf_{(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta^K}(\widehat{J}_n(K,\underline{t},\underline{\theta}) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)) \leq \frac{n}{v_n}(K^*-K).\Big). \end{aligned} \tag{2.26}$$

i-) For $K < K^*$, we decompose $\widehat{J}_n(K,\underline{t},\underline{\theta}) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*) = n(d_n(\underline{t},\underline{\theta}) + e_n(\underline{t},\underline{\theta}))$ where e_n is defined in Lemma 2.6.5 and

$$d_n(\underline{t},\underline{\theta}) = 2\left[\sum_{j=1}^{K^*}\frac{n_j^*}{n}\Big(\frac{\widehat{L}_n(T_j^*,\theta_j^*)}{n_j^*} - \mathcal{L}_j(\theta_j^*)\Big) + \sum_{k=1}^{K}\sum_{j=1}^{K^*}\frac{n_{kj}}{n}\Big(\mathcal{L}_j(\theta_k) - \frac{\widehat{L}_n(T_j^*\cap T_k,\theta_k)}{n_{kj}}\Big)\right].$$

It comes from the relation (2.26) that :

$$P(\widehat{K}_n = K) \leq P\Big(\inf_{(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta_K}(d_n(\underline{t},\underline{\theta}) + e_n(\underline{t},\underline{\theta})) \leq \frac{1}{v_n}(K^*-K)\Big). \tag{2.27}$$

Corollary 2.6.1 ensures that $d_n(\underline{t},\underline{\theta})) \to 0$ a.s. and uniformly on $\mathcal{F}_K\times\Theta^K$. By Lemma 2.6.5, there exists $C_K > 0$ such that $e_n(\underline{t},\underline{\theta}) \geq C_K\|\underline{t}-\underline{t}^*\|_\infty/n$ for all $(\underline{t},\underline{\theta}) \in \mathcal{F}_K\times\Theta^K$. But, since $K < K^*$, for any $\underline{t}\in\mathcal{F}_K$, we have $\|\underline{t}-\underline{t}^*\|_\infty/n = \|\underline{\tau}-\underline{\tau}^*\|_\infty \geq \min_{1\leq j\leq K^*}(\tau_j^*-\tau_{j-1}^*)/2$ that is positive by assumption. Then $e_n(\underline{t},\underline{\theta}) > 0$ for all $(\underline{t},\underline{\theta}) \in \mathcal{F}_K\times\Theta^K$ and since $1/v_n \underset{n\to\infty}{\longrightarrow} 0$, we deduce from (2.27) that $P(\widehat{K}_n = K) \underset{n\to\infty}{\longrightarrow} 0$.

ii-) Now let $K^* < K \leq K_{\max}$. From (2.27) and the Markov Inequality we have :

$$\begin{aligned} P(\widehat{K}_n = K) &\leq P\Big(\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*) + \frac{n}{v_n}(K-K^*) \leq 0\Big) \\ &\leq P\Big(|\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)| \geq \frac{n}{v_n}\Big) \\ &\leq \frac{v_n}{n}\mathbb{E}|\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)|. \end{aligned} \tag{2.28}$$

Denote $\widehat{\underline{t}}_n = (\widehat{t}_{n,1},\cdots,\widehat{t}_{n,K})$. By Lemma 2.6.4, there exists some subset $\{k_j\ ,1\leq j\leq K^*-1\}$ of $\{1,\cdots,K-1\}$ such that for any $j=1,\cdots,K^*-1$, $\widehat{t}_{n,k_j}/n \to \tau_j^*$. Denoting $k_0 = 0$ and $k_{K^*} = K$, we have :

$$\begin{aligned} \widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*) &= 2\Big(\sum_{j=1}^{K^*}\widehat{L}_n(T_j^*,\theta_j^*) - \sum_{k=1}^{K}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big) \\ &= 2\sum_{j=1}^{K^*}\Big[\widehat{L}_n(T_j^*,\theta_j^*) - \sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big] \end{aligned}$$

and from (2.28) we deduce that :

$$\begin{aligned} P(\widehat{K}_n = K) &\leq \frac{2v_n}{n}\sum_{j=1}^{K^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*,\theta_j^*) - \sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big| \\ &\leq C\sum_{j=1}^{K^*}\frac{v_{n_j^*}}{n_j^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*,\theta_j^*) - \sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big|. \end{aligned}$$

Since for any $j = 1, \cdots, K^* - 1$, it comes easily from the proof of Lemma 2.6.3 that

$$\frac{v_{n_j^*}}{n_j^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*, \theta_j^*) - \sum_{k=k_{j-1}+1}^{k_j} \widehat{L}_n(\widehat{T}_{n,k}, \widehat{\theta}_{n,k})\Big| \underset{n\to\infty}{\longrightarrow} 0,$$

and therefore $\mathrm{P}(\widehat{K}_n = K) \underset{n\to\infty}{\longrightarrow} 0$. $\square$

Proof of Lemma 2.6.5 Let $K \geq 1$ and consider the real function υ define on $\Theta \times \Theta$ by :

$$\upsilon(\theta, \theta') = \begin{cases} \min\limits_{1\leq j\leq K^*}[\max(\mathcal{L}_j(\theta_j^*) - \mathcal{L}_j(\theta),\ \mathcal{L}_j(\theta_j^*) - \mathcal{L}_j(\theta'))] & \text{if} \ \ \theta \neq \theta' \\ 0 \ \ \text{if} \ \ \theta = \theta'. \end{cases}$$

The function υ has positive values and $\upsilon(\theta, \theta') = 0$ if and only if $\theta = \theta'$ since the function $\theta \mapsto \mathcal{L}_j(\theta)$ has a strict maximum in θ_j^* (see [38]). By Lemma 3.3 of [50], there exists $C_{\theta^*} > 0$ such that for any $(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta_K$

$$\sum_{j=1}^{K^*}\sum_{k=1}^{K} \frac{n_{kj}}{n}\upsilon(\theta_k, \theta_j^*) \geq \frac{C_{\theta^*}}{n}\|\underline{t} - \underline{t}^*\|_\infty.$$

Moreover, for any $j = 1, \cdots, K^*$ and $\theta \in \Theta$, $\mathcal{L}_j(\theta_j^*) - \mathcal{L}_j(\theta) \geq \upsilon(\theta, \theta_j^*)$ and denoting $C_K = 2C_{\theta^*}$ the result follows immediately. $\square$

2.6.6 Proof of Theorem 2.3.2

Assume with no loss of generality that $K^* = 2$. Denote $(u_n)_n$ a sequence satisfying $u_n \underset{n\to\infty}{\longrightarrow} \infty$, $u_n/n \underset{n\to\infty}{\longrightarrow} 0$ and $\mathbb{P}(|\widehat{\underline{t}}_n - \underline{t}^*| > u_n) \underset{n\to\infty}{\longrightarrow} 0$ (for example $u_n = n\sqrt{\max(\mathbb{E}|\widehat{\tau}_n - \tau^*|, n^{-1})}$). For $\delta > 0$, as we have

$$\mathbb{P}(|\widehat{\underline{t}}_n - \underline{t}^*| > \delta) \leq \mathbb{P}(\delta < |\widehat{\underline{t}}_n - \underline{t}^*| \leq u_n) + \mathbb{P}(|\widehat{\underline{t}}_n - \underline{t}^*|_m > u_n)$$

it suffices to show that $\lim\limits_{\delta\to\infty}\lim\limits_{n\to\infty}\mathbb{P}(\delta < |\widehat{t}_n - t^*| \leq u_n) = 0$.

Denote $V_{\delta,u_n} = \{\ t \in \mathbb{Z}/\ \ \delta < |t - t^*| \leq u_n\ \}$. Then,

$$\mathbb{P}(\delta < |\widehat{t}_n - t^*| \leq u_n) \leq \mathbb{P}\Big(\min_{t\in V_{\delta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \leq 0\Big).$$

Let $t \in V_{\delta,u_n}$ (for example $t \geq t^*$). With the notation of the proof of Theorem 2.3.1, we have $\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*)) \geq \widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1))$ and from (2.22) we obtain :

$$\frac{\widehat{I}_n(t) - \widehat{I}_n(t^*)}{t - t^*} \geq \frac{2}{t - t^*}\Big(\widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2))\Big).$$

We conclude in two steps :

i-) We show that $\dfrac{1}{t - t^*}\Big(\widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_1))\Big) > 0$ for n large enough.

Then $\dfrac{\widehat{L}_n(T_1, \theta)}{n} = \dfrac{t^*}{n}\dfrac{\widehat{L}_n(T_1^*, \theta)}{t^*} + \dfrac{t - t^*}{n}\dfrac{\widehat{L}_n(T_1 \cap T_2^*, \theta)}{t - t^*}$ and since $\dfrac{t - t^*}{n} \leq \dfrac{u_n}{n} \underset{n\to\infty}{\longrightarrow} 0$ and

$$\widehat{\theta}_n(T_1) = \underset{\theta\in\Theta}{\mathrm{Argmax}}\left(\frac{1}{n}\widehat{L}_n(T_1, \theta)\right) \ \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}} \ \theta_1^*.$$

It comes that $\dfrac{1}{t - t^*}\Big(\widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_1))\Big)$ converges a.s. and uniformly on V_{δ,u_n} to $\mathcal{L}_2(\theta_2^*) - \mathcal{L}_2(\theta_1^*) > 0$.

ii-) We show that $\dfrac{1}{t-t^*}\Big(\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2,\widehat{\theta}_n(T_2))\Big) \xrightarrow[n,\delta\to\infty]{a.s.} 0$. For large value of n, we remark that $\widehat{\theta}_n(T_2) \in \overset{\circ}{\Theta}$ so that $\partial\widehat{L}_n(T_2,\widehat{\theta}_n(T_2))/\partial\theta = 0$. The mean value theorem on $\partial\widehat{L}_n/\partial\theta_i$ for any $i=1,\dots,d$ gives the existence of $\widetilde{\theta}_{n,i} \in [\widehat{\theta}_n(T_2),\widehat{\theta}_n(T_2^*)]$ such that :

$$0 \;=\; \frac{\partial\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))}{\partial\theta_i} + \frac{\partial^2\widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\widehat{\theta}_n(T_2)) - \widehat{\theta}_n(T_2^*)) \tag{2.29}$$

where for $a,b \in \mathbb{R}^d$, $[a,b] = \{(1-\lambda)a+\lambda b \;;\; \lambda \in [0,1]\}$. Using the equalities $\widehat{L}_n(T_2^*,\theta) = \widehat{L}_n(T_1\cap T_2^*,\theta) + \widehat{L}_n(T_2,\theta)$ and $\partial\widehat{L}_n(T_2^*,\widehat{\theta}_n(T_2^*))/\partial\theta = 0$, it comes from (2.29) :

$$\frac{\partial\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta_i} = \frac{\partial^2\widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\widehat{\theta}_n(T_2)) - \widehat{\theta}_n(T_2^*)), \qquad \forall i=1,\dots,d,$$

and it follows :

$$\frac{1}{t-t^*}\frac{\partial\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta} = \frac{n-t}{t-t^*}A_n\cdot(\widehat{\theta}_n(T_2) - \widehat{\theta}_n(T_2^*)) \tag{2.30}$$

with $A_n := \Big(\dfrac{1}{n-t}\dfrac{\partial^2\widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}\Big)_{1\le i\le d}$. Corollary 2.6.1 ii-) gives that :

$$\frac{1}{t-t^*}\frac{\partial\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta} \xrightarrow[n,\delta\to\infty]{a.s.} \frac{\partial\mathcal{L}_2(\theta_2^*)}{\partial\theta} = 0$$

and $A_n \xrightarrow[n,\delta\to\infty]{a.s.} -\dfrac{1}{2}\mathbb{E}\Big(\dfrac{\partial^2 q_{0,2}(\theta_2^*)}{\partial\theta^2}\Big)$. Under assumption (Var), $\mathbb{E}\Big(\dfrac{\partial^2 q_{0,2}(\theta_2^*)}{\partial\theta^2}\Big)$ is a nonsingular matrix (see [8]). Then, we deduce from (2.30) that

$$\frac{n-t}{t-t^*}(\widehat{\theta}_n(T_2) - \widehat{\theta}_n(T_2^*)) \xrightarrow[n,\delta\to\infty]{a.s.} 0. \tag{2.31}$$

We conclude by the Taylor expansion on $\widehat{L}_n$ that gives

$$\begin{aligned}\frac{1}{t-t^*}&|\widehat{L}_n(T_2,\widehat{\theta}_n(T_2)) - \widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))| \\ &\le \frac{1}{2(t-t^*)}\|\widehat{\theta}_n(T_2)) - \widehat{\theta}_n(T_2^*)\|^2 \sup_{\theta\in\Theta}\Big\|\frac{\partial^2\widehat{L}_n(T_2,\theta)}{\partial\theta^2}\Big\| \to 0 \quad \text{a.s.} \;\square\end{aligned}$$

2.6.7 Proof of Theorem 2.3.3

First, $(\widehat{\theta}_n(\widehat{T}_j) - \theta_j^*) = (\widehat{\theta}_n(\widehat{T}_j) - \widehat{\theta}_n(T_j^*)) + (\widehat{\theta}_n(T_j^*) - \theta_j^*)$ for any $j \in \{1,\cdots,K^*\}$. By Theorem 2.3.2 it comes $\widehat{t}_j - t_j^* = o_P(\log(n))$. Using relation (2.31), we obtain : $\widehat{\theta}_n(\widehat{T}_j) - \widehat{\theta}_n(T_j^*) = o_P(\dfrac{\log(n)}{n})$. Hence, $\sqrt{n_j^*}(\widehat{\theta}_n(\widehat{T}_j) - \widehat{\theta}_n(T_j^*)) \xrightarrow[n\to\infty]{P} 0$ and it suffices to show that $\sqrt{n_j^*}(\widehat{\theta}_n(T_j^*) - \theta_j^*) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d(0, F(\theta_j^*)^{-1}G(\theta_j^*)F(\theta_j^*)^{-1})$ to conclude.
For large value of n, $\widehat{\theta}_n(T_j^*) \in \overset{\circ}{\Theta}$. By the mean value theorem, there exists $(\widetilde{\theta}_{n,k})_{1\le k\le d} \in [\widehat{\theta}_n(T_j^*),\theta_j^*]$ such that

$$\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta_k} = \frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta_k} + \frac{\partial^2 L_n(T_j^*,\widetilde{\theta}_{n,k})}{\partial\theta\partial\theta_k}(\widehat{\theta}_n(T_j^*) - \theta_j^*). \tag{2.32}$$

Let $F_n = -2\Big(\frac{1}{n_j^*}\frac{\partial^2 L_n(T_j^*,\widetilde{\theta}_{n,k})}{\partial\theta\partial\theta_k}\Big)_{1\leq k\leq d}$. By Lemma 2.6.3 and Corollary 2.6.1, $F_n \xrightarrow[n\to\infty]{\text{a.s.}} F(\theta_j^*)$ (where $F(\theta_j^*)$ is defined by (2.9)). But, under (Var), $F(\theta_j^*)$ is a non singular matrix (see [8]). Thus, for n large enough, F_n is invertible and (2.32) gives

$$\sqrt{n_j^*}(\widehat{\theta}_n(T_j^*)-\theta_j^*) = -2F_n^{-1}\Big[\frac{1}{\sqrt{n_j^*}}\Big(\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}-\frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta}\Big)\Big].$$

As in proof of Lemma 3 of [8], it is now easy to show that :

$$\frac{1}{\sqrt{n_j^*}}\frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta}\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}_d(0,G(\theta_j^*))$$

where $G(\theta_j^*)$ is given by (2.9). Thus, since $\partial\widehat{L}_n(T_j^*,\widehat{\theta}_n(T_j^*))/\partial\theta=0$, we have :

$$\frac{1}{\sqrt{n_j^*}}\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}=\frac{1}{\sqrt{n_j^*}}\Big(\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}-\frac{\partial\widehat{L}_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}\Big)\xrightarrow[n\to\infty]{\text{a.s.}}0.$$

We conclude using Lemma 2.6.3 and the fact that $1/\sqrt{n}=O(v_n/n)$. $\square$

Chapitre 3

Testing for parameter constancy in general causal time series models

Abstract

We consider a process $X = (X_t)_{t\in\mathbb{Z}}$ belonging to a large class of causal models including AR(∞), ARCH(∞), TARCH(∞),... processes. We assume that the model depends on a parameter $\theta_0 \in \Theta \subset \mathbb{R}^d$ and consider the problem of testing for change in the parameter. Two statistics $\widehat{Q}_n^{(1)}$ and $\widehat{Q}_n^{(2)}$ are constructed using quasi-likelihood estimator (QLME) of the parameter. Under the null hypothesis that there is no change, it is shown that each of these two statistics weakly converges to the supremum of the sum of the squares of independent Brownian bridges. Under the alternative of a change in the parameter, we show that the test statistic $\widehat{Q}_n = \max(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)})$ diverges to infinity. Some simulation results for AR(1), ARCH(1), GARCH(1,1) and TARCH(1) models are reported to show the applicability and the performance of our procedure with comparisons to some other approaches.

Keywords : Semi-parametric test ; Change of parameters ; Causal processes ; Quasi-maximum likelihood estimator ; Weak convergence.

Note

The content of this chapter is based on a paper, published in the Journal of Time Series Analysis.

3.1 Introduction

Many statistical data can be represented by models which may change over time, for instance hydraulic flow, climate data. Before any inference on these data, it is crucial to test whether a change has not occurred in the model.

Since Page [61] in 1955, real advances have been done about tests for change detection. Horvath [32] proposed a test for detecting a change in the parameter of autoregressive processes based on weighted supremum and L_p-functionals of the residual sums. The CUSUM statistic which was successfully applied by Brown *et al.* [21] in 1975, was extended by Inclán and Tiao [37] for detecting multiple changes in variance of independent random variables. Numerous works devoted to the CUSUM-type procedure, for instance Kim *et al.* [42] for testing change in parameters of GARCH(1,1), Kokoszka and Leipus [44] in the specific case of ARCH(∞) or Aue *et al.* [5] for testing breaks in covariance. Kulperger and Yu [48] studied the high moment partial sum process based on residuals and applied it to the residual CUSUM test in GARCH model. Horváth *et al.* [33] suggested to compute the ratio of the CUSUM functionals instead of the differences for testing change in the mean of a time series. Berkes *et al.* [14] used a test based on approximate likelihood scores for testing parameter constancy in GARCH(p,q) models. Lee and Na [54] proposed a test based on conditional least-squares estimator. The procedure is numerically simple but requires high moment assumptions (for example : moment of order 8 for ARCH models). Lee and Song [55] developed a test based on quasi-maximum likelihood estimator for parameter change in ARMA-GARCH models. The procedure does not take into account the change-point alternative ; so the consistency in power is not ensured. The present work is a new contribution to the challenging problem of test for change detection.

In this paper, we consider a general class $\mathcal{M}_T(M, f)$ of causal (non-anticipative) time series. Let $M, f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be measurable functions, $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) random variables called the innovations and satisfying $\mathrm{var}(\xi_0) = \sigma^2$ and Θ a compact subset of $\mathbb{R}^d$. Let $T \subset \mathbb{Z}$, and for any $\theta \in \Theta$, define

Class $\mathcal{M}_T(M_\theta, f_\theta)$ **:** *The process* $X = (X_t)_{t\in\mathbb{Z}}$ *belongs to* $\mathcal{M}_T(M_\theta, f_\theta)$ *if it satisfies the relation :*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{3.1}$$

The existence and properties of this general class of affine processes were studied in Bardet and Wintenberger [8]. Numerous classical time series are included in $\mathcal{M}_{\mathbb{Z}}(M, f)$: for instance AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes.

Now, assume that a trajectory $(X_1, \cdots, X_n)$ of $X = (X_t)_{t\in\mathbb{Z}}$ is observed and consider the following hypothesis :

$\mathbf{H}_0$: there exists $\theta_0 \in \Theta$ such that $(X_1, \cdots, X_n)$ belongs to the class $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0}, f_{\theta_0})$;
$\mathbf{H}_1$: there exist $K \geq 2$, $\theta_1^*, \cdots, \theta_K^* \in \Theta$ with $\theta_j^* \neq \theta_{j+1}^*$, such that $(X_1, \cdots, X_n)$ belongs to $\bigcap_{j=1}^{K} \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*})$ where $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \cdots, t_j^*\}$ with $0 = t_0^* < t_1^* < \cdots < t_{K-1}^* < t_K^* = n$.

Thus, it is easy to see that under H_1 the property of stationary is lost after the first change. This is not the case in many existing works (for instance Kouamo *et al.* [46]) where the stationarity or the K-th order stationarity after the change is an essential assumption.

In this paper we study a new test for change detection (see Bardet *et al.* [9] for the procedure of the estimation of the instants of change). We consider a semi-parametric test statistic based on the QLME which is a modification of the statistic proposed by Lee *et al.* [53]. For $k, k' \in \{1, \cdots, n-1\}$ (with $k \leq k'$) let $\widehat{\theta}_n(X_k, \cdots, X_{k'})$ be the QLME of the parameter computed on $\{k, \cdots, k'\}$. The basic idea of our procedure is that : under H_0, $\widehat{\theta}_n(X_1, \cdots, X_k)$ and $\widehat{\theta}_n(X_{k+1}, \cdots, X_n)$ are close to $\widehat{\theta}_n(X_1, \cdots, X_n)$ and the distances $\|\widehat{\theta}_n(X_1, \cdots, X_k) - \widehat{\theta}_n(X_1, \cdots, X_n)\|$ and $\|\widehat{\theta}_n(X_{k+1}, \cdots, X_n) - \widehat{\theta}_n(X_1, \cdots, X_n)\|$ are not too large. Thus, we show that the test statistic is finite under the null hypothesis and diverges to infinity under the alternative of change in the parameter of model. Simulation results compared to some other tests show that our procedure provides satisfactory results in any case. In Section 2 we present assumptions, some examples and the construction of the test statistic. In Section 3 we give some asymptotic results. The empirical studies of AR(1), ARCH(1), GARCH(1,1) and TARCH(1) are detailed in Section 4 and the proofs of the main results are presented in Section 5.

3.2 Assumptions and test statistics

3.2.1 Assumptions on the class of models $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$

Let $\theta \in \mathbb{R}^d$ and M_θ and f_θ be numerical functions such that for all $(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $M_\theta\big((x_i)_{i\in\mathbb{N}}\big) \neq 0$ and $f_\theta\big((x_i)_{i\in\mathbb{N}}\big) \in \mathbb{R}$. Denote $h_\theta := M_\theta^2$. We will use the following norms :

1. $\|\cdot\|$ applied to a vector denotes the Euclidean norm of the vector ;
2. for any compact set $\Theta \subseteq \mathbb{R}^d$ and for any $g : \Theta \longrightarrow \mathbb{R}^{d'}$, $\|g\|_\Theta = \sup_{\theta\in\Theta}(\|g(\theta)\|)$.

Throughout the sequel, we will assume that the functions $\theta \mapsto M_\theta$ and $\theta \mapsto f_\theta$ are twice continuously differentiable on Θ. Let $\Psi_\theta = f_\theta$, M_θ and $i = 0$, 1, 2, then define

Assumption $\mathbf{A}_i(\Psi_\theta, \Theta)$: *Assume that $\|\partial^i \Psi_\theta(0)/\partial\theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real number $(\alpha_i^{(k)}(\Psi_\theta, \Theta))_{i\geq 1}$ such that $\sum\limits_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \Theta) < \infty$ satisfying*

$$\Big\|\frac{\partial^i \Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i \Psi_\theta(y)}{\partial\theta^i}\Big\|_\Theta \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \Theta)|x_k - y_k| \quad \text{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

In the sequel we refer to the particular case called "ARCH-type process" if $f_\theta = 0$ and if the following assumption holds with $h_\theta = M_\theta^2$:

Assumption $\mathbf{A}_i(h_\theta, \Theta)$: *Assume that $\|\partial^i h_\theta(0)/\partial\theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real number $(\alpha_i^{(k)}(h_\theta, \Theta))_{i\geq 1}$ such as $\sum\limits_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \Theta) < \infty$ satisfying*

$$\Big\|\frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i}\Big\|_\Theta \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \Theta)|x_k^2 - y_k^2| \quad \text{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

Then define the set :

$$\Theta(r) := \{\theta \in \Theta,\ A_0(f_\theta, \{\theta\}) \text{ and } A_0(M_\theta, \{\theta\}) \text{ hold with } \sum_{k\geq 1} \alpha_k^{(0)}(f_\theta, \theta) + (\mathbb{E}|\xi_0|^r)^{1/r} \sum_{k\geq 1} \alpha_k^{(0)}(M_\theta, \theta) < 1\}$$
$$\cup \{\theta \in \Theta,\ f_\theta = 0 \text{ and } A_0(h_\theta, \{\theta\}) \text{ hold with } (\mathbb{E}|\xi_0|^r)^{2/r} \sum_{k\geq 1} \alpha_k^{(0)}(h_\theta, \theta) < 1\}.$$

The Lipschitz-type hypothesis $A_i(\Psi_\theta,\Theta)$ are classical when studying the existence of solutions of the general model. If $\theta \in \Theta(r)$ the existence of a unique causal, stationary and ergodic solution $X = (X_t)_{t\in\mathbb{Z}} \in \mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$ is ensured (see [8]). The subset $\Theta(r)$ is defined as a reunion to consider accurately general causal models and ARCH-type models simultaneously.
The following assumptions are needed to study QLME property.
Assumption D(Θ) **:** $\exists \underline{h} > 0$ such that $\inf\limits_{\theta\in\Theta}(|h_\theta(x)|) \geq \underline{h}$ for all $x \in \mathbb{R}^{\mathbb{N}}$.

Assumption Id(Θ**) :** For all $(\theta,\theta') \in \Theta^2$,

$$\Big(f_\theta(X_0,X_{-1},\cdots) = f_{\theta'}(X_0,X_{-1},\cdots) \text{ and } h_\theta(X_0,X_{-1},\cdots) = h_{\theta'}(X_0,X_{-1},\cdots) \text{ a.s.}\Big) \Rightarrow \ \theta = \theta'.$$

Assumption Var(Θ**) :** For all $\theta \in \Theta$, one of the families $\big(\dfrac{\partial f_\theta}{\partial \theta^i}(X_0,X_{-1},\cdots)\big)_{1\leq i\leq d}$ or $\big(\dfrac{\partial h_\theta}{\partial \theta^i}(X_0,X_{-1},\cdots)\big)_{1\leq i\leq d}$ is a.s. linearly independent.

As in [8], we will make the convention that if $\mathbf{A}_i(M_\theta,\Theta)$ holds then $\alpha_\ell^{(i)}(h_\theta,\Theta) = 0$ and if $\mathbf{A}_i(h_\theta,\Theta)$ holds then $\alpha_\ell^{(i)}(M_\theta,\Theta) = 0$. Denote :

Assumption K$(f_\theta, M_\theta, \Theta)$: for i= 0, 1, 2, $\mathbf{A}_i(f_\theta,\Theta)$ and $\mathbf{A}_i(M_\theta,\Theta)$ (or $\mathbf{A}_i(h_\theta,\Theta)$) hold and there exists $\ell > 2$ such that $\alpha_j^{(i)}(f_\theta,\Theta) + \alpha_j^{(i)}(M_\theta,\Theta) + \alpha_j^{(i)}(h_\theta,\Theta) = \mathcal{O}(j^{-\ell})$, for i= 0, 1.

3.2.2 Examples

1. AR(∞) models.
   Consider the AR(∞) process defined by :
   $$X_t = \sum_{k\geq 1} \phi_k(\theta_0^*) X_{t-k} + \xi_t \ , \ t \in \mathbb{Z}$$
   with $\theta_0^* \in \Theta$, where Θ is a compact subset of $\mathbb{R}^d$ such that $\sum_{k\geq 1} \|\phi_k(\theta)\|_\Theta < 1$. The process belongs to the class $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ where $f_\theta(x_1,\cdots) = \sum_{k\geq 1}\phi_k(\theta)x_k$ and $M_\theta \equiv 1$ for all $\theta \in \Theta$. Then Assumptions D(Θ) and $A_0(f_\theta,\Theta)$ hold with $\underline{h} = 1$ and $\alpha_k^{(0)}(f_\theta,\Theta) = \|\phi_k(\theta)\|_\Theta$. If there exists $\ell > 2$ and ϕ_k twice differentiable such as $\|\phi_k(\theta)\|_\Theta = \|\phi_k'(\theta)\|_\Theta = \|\phi_k''(\theta)\|_\Theta = O(k^{-\ell})$, then Assumptions $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds. Moreover, if ξ_0 is a nondegenerate random variable, Id(Θ) and Var(Θ) hold. For any $r \geq 1$ such that $\mathbb{E}|\xi_0|^r < \infty$, $\Theta(r) = \Theta$.
2. GARCH(p,q) models.
   Consider the GARCH(p,q) process defined by :
   $$X_t = \sigma_t\,\xi_t \ , \ \sigma_t^2 = \alpha_0^* + \sum_{k=1}^{q} \alpha_k^* X_{t-k}^2 + \sum_{k=1}^{p} \beta_k^* \sigma_{t-k}^2 \ , \quad t \in \mathbb{Z}$$
   with $\mathbb{E}(\xi_0^2) = 1$ and $\theta_0^* := (\alpha_0^*,\cdots,\alpha_q^*,\beta_1^*,\cdots,\beta_p^*) \in \Theta$ where Θ is a compact subset of $]0,\infty[\times[0,\infty[^{p+q}$ such that $\sum_{k=1}^q \alpha_k + \sum_{k=1}^p \beta_k < 1$ for all $\theta \in \Theta$. Then there exists (see Bollerslev [20] or Nelson and Cao [60]) a nonnegative sequence $(\psi_k(\theta_0^*))_{k\geq 0}$ such that $\sigma_t^2 = \psi_0(\theta_0^*) + \sum_{k\geq 1}\psi_k(\theta_0^*)X_{t-k}^2$ with $\psi_0(\theta_0^*) = \alpha_0^*/(1-\sum_{k=1}^p \beta_k^*)$. This process belongs to a class $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ where $M_\theta(x_1,\cdots) = \sqrt{\psi_0(\theta) + \sum_{k\geq 1}\psi_k(\theta)x_k}$ and

$f_\theta \equiv 0\ \forall \theta \in \Theta$. Assumptions D($\Theta$) holds with $\underline{h} = \inf_{\theta\in\Theta}(\alpha_0)$. If there exists $0 < \rho_0 < 1$ such that for any $\theta \in \Theta$, $\sum_{k=1}^q \alpha_k + \sum_{k=1}^p \beta_k \leq \rho_0$ then the sequences $(\|\psi_k(\theta)\|_\Theta)_{k\geq 1}$, $(\|\psi'_k(\theta)\|_\Theta)_{k\geq 1}$ and $(\|\psi''_k(\theta)\|_\Theta)_{k\geq 1}$ decay exponentially fast (see Berkes *et al.* [13]), thus Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds. Moreover, if ξ_0^2 is a nondegenerate random variable, Id(Θ) and Var(Θ) hold. For $r \geq 2$ denote

$$\Theta(r) = \{\theta \in \Theta \ ; \ (\mathbb{E}|\xi_0|^r)^{2/r} \sum_{k=1}^q \alpha_k + \sum_{k=1}^p \beta_k < 1\}.$$

3. TARCH(∞) model.
Consider a Threshold ARCH(∞) model (introduced by Rabemananjara and Zakoïan [62]) defined by :

$$X_t = \sigma_t \xi_t \ , \ \sigma_t = b_0(\theta_0^*) + \sum_{k\geq 1} \Big(b_k^+(\theta_0^*) \max(X_{t-k}, 0) - b_k^-(\theta_0^*) \min(X_{t-k}, 0) \Big), \ t \in \mathbb{Z}$$

with $\theta_0^* \in \Theta$, where Θ is a compact subset of $\mathbb{R}^d$ satisfying $\sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < \infty$. Then $f_\theta \equiv 0$ and $(A_0(M_\theta, \Theta))$ holds with $\alpha_k^{(0)}(M_\theta, \Theta) = \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta)$. Assume that for any $\theta \in \Theta$, $b_0(\theta)$, $b_k^+(\theta)$ and $b_k^-(\theta)$ are non negative real numbers. If $\inf_{\theta\in\Theta}(b_0(\theta)) > 0$ then D(Θ) holds. Assume the functions $\theta \mapsto b_k^+(\theta)$ and $\theta \mapsto b_k^-(\theta)$ are twice differentiable and there exists $\ell > 2$ such that $\|b_k^+(\theta)\|_\Theta = \|\partial b_k^+(\theta)/\partial\theta\|_\Theta = \|\partial^2 b_k^+(\theta)/\partial\theta^2\|_\Theta = O(k^{-\ell})$ (the same for b_k^-). Then $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds. Id(Θ) and Var(Θ) hold if ξ_0 is a nondegenerate random variable. For $r \geq 1$, denote

$$\Theta(r) = \{\theta \in \Theta \ ; \ (\mathbb{E}|\xi_0|^r)^{1/r} \sum_{k\geq 1} \max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < 1\}.$$

3.2.3 Test statistics

Assume that a trajectory $(X_1, \cdots, X_n)$ is observed. If $(X_1, \cdots, X_n) \in \mathcal{M}_{\{1,\cdots,n\}}(M_\theta, f_\theta)$, then for $T \subset \{1, \cdots, n\}$, the conditional quasi-(log)likelihood computed on T is given by :

$$L_n(T, \theta) := -\frac{1}{2} \sum_{t\in T} q_t(\theta) \ \text{ with } \ q_t(\theta) = \frac{(X_t - f_\theta^t)^2}{h_\theta^t} + \log(h_\theta^t)$$

where $f_\theta^t = f_\theta(X_{t-1}, X_{t-2} \ldots)$, $M_\theta^t = M_\theta(X_{t-1}, X_{t-2} \ldots)$ and $h_\theta^t = M_\theta^{t\,2}$. As is was now usually done (see [8]), we approximate this conditional log-likelihood by :

$$\widehat{L}_n(T, \theta) := -\frac{1}{2} \sum_{t\in T} \widehat{q}_t(\theta) \quad \text{with} \quad \widehat{q}_t(\theta) := \frac{\big(X_t - \widehat{f}_\theta^t\big)^2}{\widehat{h}_\theta^t} + \log\big(\widehat{h}_\theta^t\big)$$

where $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \ldots, X_1, 0, 0, \cdots)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \ldots, X_1, 0, 0, \cdots)$ and $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$. For $T \subset \{1, \cdots, n\}$, define the quasi-likelihood estimator computed on T by $\widehat{\theta}_n(T) := \operatorname{argmax}_{\theta\in\Theta}(\widehat{L}_n(T, \theta))$. Now, for $T \subset \{1, \cdots, n\}$ define

$$\widehat{G}_n(T) := \frac{1}{Card(T)} \sum_{t\in T} \Big(\frac{\partial \widehat{q}_t(\widehat{\theta}_n(T))}{\partial\theta}\Big)\Big(\frac{\partial \widehat{q}_t(\widehat{\theta}_n(T))}{\partial\theta}\Big)'$$

and

$$\widehat{F}_n(T) := -\frac{2}{Card(T)} \Big(\frac{\partial^2 \widehat{L}_n(T, \widehat{\theta}_n(T))}{\partial\theta\partial\theta'}\Big) = \frac{1}{Card(T)} \sum_{t\in T} \frac{\partial^2 \widehat{q}_t(\widehat{\theta}_n(T))}{\partial\theta\partial\theta'}.$$

The matrix $\widehat{G}_n(T)$ is symmetric positive semi-definite. For $k = 1, \cdots, n-1$, denote $T_k = \{1, \cdots, k\}$, $\overline{T}_k = \{k+1, \cdots, n\}$ and define

$$\widehat{\Sigma}_{n,k} := \frac{k}{n}\widehat{F}_n(T_k)\widehat{G}_n(T_k)^{-1}\widehat{F}_n(T_k)\mathbb{1}_{\det(\widehat{G}_n(T_k))\neq 0}+\frac{n-k}{n}\widehat{F}_n(\overline{T}_k)\widehat{G}_n(\overline{T}_k)^{-1}\widehat{F}_n(\overline{T}_k)\mathbb{1}_{\det(\widehat{G}_n(\overline{T}_k))\neq 0}.$$

For $k = 1, \cdots, n-1$, $\widehat{\Sigma}_{n,k}$ is symmetric positive semi-definite. Let $(v_n)_{n\in\mathbb{N}}$ be a sequence satisfying $v_n \to \infty$ and $v_n/n \to 0$ (as $n \to \infty$). Denote $\Pi_n = [v_n, n - v_n] \cap \mathbb{N}$ and define the statistics :

$$\widehat{Q}_n^{(1)} := \max_{k\in\Pi_n}\widehat{Q}_{n,k}^{(1)} \quad \text{where} \quad \widehat{Q}_{n,k}^{(1)} := \frac{k^2}{n}\big(\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)\big)'\widehat{\Sigma}_{n,k}\big(\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)\big);$$

$$\widehat{Q}_n^{(2)} := \max_{k\in\Pi_n}\widehat{Q}_{n,k}^{(2)} \quad \text{where} \quad \widehat{Q}_{n,k}^{(2)} := \frac{(n-k)^2}{n}\big(\widehat{\theta}_n(\overline{T}_k) - \widehat{\theta}_n(T_n)\big)'\widehat{\Sigma}_{n,k}\big(\widehat{\theta}_n(\overline{T}_k) - \widehat{\theta}_n(T_n)\big).$$

The new test statistic is defined by

$$\widehat{Q}_n := \max\big(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}\big).$$

Remark 3.2.1. The sequence (v_n) is very important in practice. A numerical algorithm can be used to compute the estimator $\widehat{\theta}_n(T)$; so, a minimum size of T is needed for the convergence of the algorithm. Notice that, the matrix $\widehat{G}_n(T)$ can be ill-conditioned, therefore if it is not numerically well approximated, its inversion may reduce the performance of the procedure. To perform the asymptotic size distortion of the test, it is assumed that (v_n) tends to infinity. Under alternative, the change-points must belong to Π_n to be detected, for this reason we assume $v_n << n$. To keep the accuracy of the procedure, it is useful to take a sequence (v_n) which does not increase too fast. We evaluated the procedure with $v_n = [\log n]$, $[(\log n)^2]$, $[(\log n)^3]$ and recommend to use $v_n = [(\log n)^2]$ for linear model and $v_n = [(\log n)^\delta]$ (with $5/2 \leq \delta < 3$) for GARCH-type and TARCH model.

Lee and Song [55] constructed a test for detecting changes in parameters of ARMA-GARCH models. The test is based on the statistic

$$\widehat{Q}_n^{(0)} := \max_{1\leq k\leq n}\Big(\frac{k^2}{n}\big(\widehat{\theta}_n(T_k)-\widehat{\theta}_n(T_n)\big)'\widehat{\Sigma}_n\big(\widehat{\theta}_n(T_k)-\widehat{\theta}_n(T_n)\big)\Big) \text{ where } \widehat{\Sigma}_n = \widehat{F}_n(T_n)\widehat{G}_n(T_n)^{-1}\widehat{F}_n(T_n). \tag{3.2}$$

Under the null hypothesis (the parameter θ_0 does not change), the estimator $\widehat{\theta}_n(T_n)$ is consistent and $\widehat{\Sigma}_n^{-1}$ is a consistent estimator of the asymptotic covariance of $\widehat{\theta}_n(T_n)$ (see [8]). Under the alternative, the model depends on several parameters and $\widehat{\theta}_n(T_n)$ may not be a consistent estimator of one of them. Therefore, the consistency of the matrix $\widehat{\Sigma}_n^{-1}$ is not ensured and the asymptotic behavior of the test statistic may be very difficult to study. Shao and Zhang (2010) pointed out that such test does not take into account the change-point alternative and can have a low power. To solve this problem, we introduce the family of matrices $\{\widehat{\Sigma}_{n,k},\ k \in \Pi_n\}$. It is easy to see that under the null hypothesis, any sequence $(\widehat{\Sigma}_{n,k_n})_{n>1,k_n\in\Pi_n}$ is consistent. On can see (in proof of Theorem 3.3.2) that under the alternative of change in the model, there exists a sequence $(\widehat{\Sigma}_{n,k_n^*})_{n>1,k_n^*\in\Pi_n}$ which is consistent.

Note that, the type of the procedure which takes the maximum of the maximum between the statistic based on the estimator computed with the observations until k $(X_1, \cdots, X_k)$ and the one computed with the observations after k $(X_{k+1}, \cdots, X_n)$ was already studied by other authors. It has been used by Berkes *et al.* [17] for a discrimination between long-range dependence and changes in mean. It was recently used by Aue *et* *al.* [3] to make inference about the specific form of the alternative (distinguishing between random walk and changes in the mean). Their procedure can give us ideas (in the future work) for distinguishing in the alternative, change in the innovation and change in the parameter of model (3.1).

3.3 Asymptotic results

3.3.1 Asymptotic behavior under the null hypothesis

Theorem 3.3.1. Assume $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$, $\mathbf{Var}$ and $\mathbf{K}(f_\theta, M_\theta, \Theta)$. Under the null hypothesis H_0 of no change, if $\theta_0 \in \overset{\circ}{\Theta}(4)$, then for $j = 1, 2$,

$$\widehat{Q}_n^{(j)} \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{0\leq\tau\leq 1} \|W_d(\tau)\|^2$$

where W_d is a d-dimensional Brownian bridge.

For any $\alpha \in (0,1)$, let C_α denote the $(1-\alpha/2)$-quantile of the distribution of $\sup\limits_{0\leq\tau\leq 1} \|W_d(\tau)\|^2$. Then, the following corollary is a straightforward application of Theorem 3.3.1 based on Bonferroni Inequality.

Corollary 3.3.1. Under assumptions of Theorem 3.3.1 :

$$\forall \alpha \in (0,1) \;\; \limsup_{n\to\infty} P\big(\widehat{Q}_n > C_\alpha\big) \leq \alpha.$$

Remark 3.3.1. The quantile values of the distribution of $\sup\limits_{0\leq\tau\leq 1} \|W_d(\tau)\|^2$ are known (see for instance Kiefer [41] for $d \in \{1, \cdots, 5\}$ or Lee *et al.* [53] for $d \in \{1, \cdots, 10\}$).

Theorem 3.3.1 and Corollary 3.3.1 imply that a large value of $\widehat{Q}_n$ means a change in the model. At a nominal level α, the critical region of the test is $(\widehat{Q}_n > C_\alpha)$.

Figure 3.1 is an illustration of the test procedure for AR(1) process. At a level $\alpha = 0.05$, for $d = 1$, $C_\alpha \simeq 2.20$. Figure 1 c-) and d-) show that, the values of $\widehat{Q}_{n,k}^{(1)}$ and $\widehat{Q}_{n,k}^{(2)}$ are all below the horizontal line which represents the limit of the critical region. Figure 3.1 e-) and f-) show that $\widehat{Q}_{n,k}^{(1)}$ and $\widehat{Q}_{n,k}^{(2)}$ are larger and increase around the point where the change occurs.

As it can be observed on Figure 3.1 and Figure 3.2 , the statistics $\widehat{Q}_n^{(1)}$ and $\widehat{Q}_n^{(2)}$ are clearly not equal. Figure 2 shows the typical example for $ARCH(1)$ with one change where $\widehat{Q}_n^{(1)} < C_\alpha$ and $\widehat{Q}_n^{(2)} > C_\alpha$. In general, we do not know if under the alternative hypothesis each of statistics $\widehat{Q}_n^{(1)}$ and $\widehat{Q}_n^{(2)}$ take large values. But we will show that their maximum diverges to infinity (see Theorem 3.3.2). This is the reason why we define the critical region as $\{\max(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}) > C_\alpha\}$.

3.3.2 The asymptotic under the alternative

In this subsection, we consider the alternative that there is a change in the model. Recall the alternative
H_1 : there exist $K \geq 2$, $\theta_1^*, \cdots, \theta_K^* \in \Theta$ with $\theta_j^* \neq \theta_{j+1}^*$, such that $(X_1, \cdots, X_n)$ belongs to $\bigcap\limits_{j=1}^{K} \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*})$ where $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \cdots, t_j^*\}$ with $0 = t_0^* < t_1^* < \cdots <$

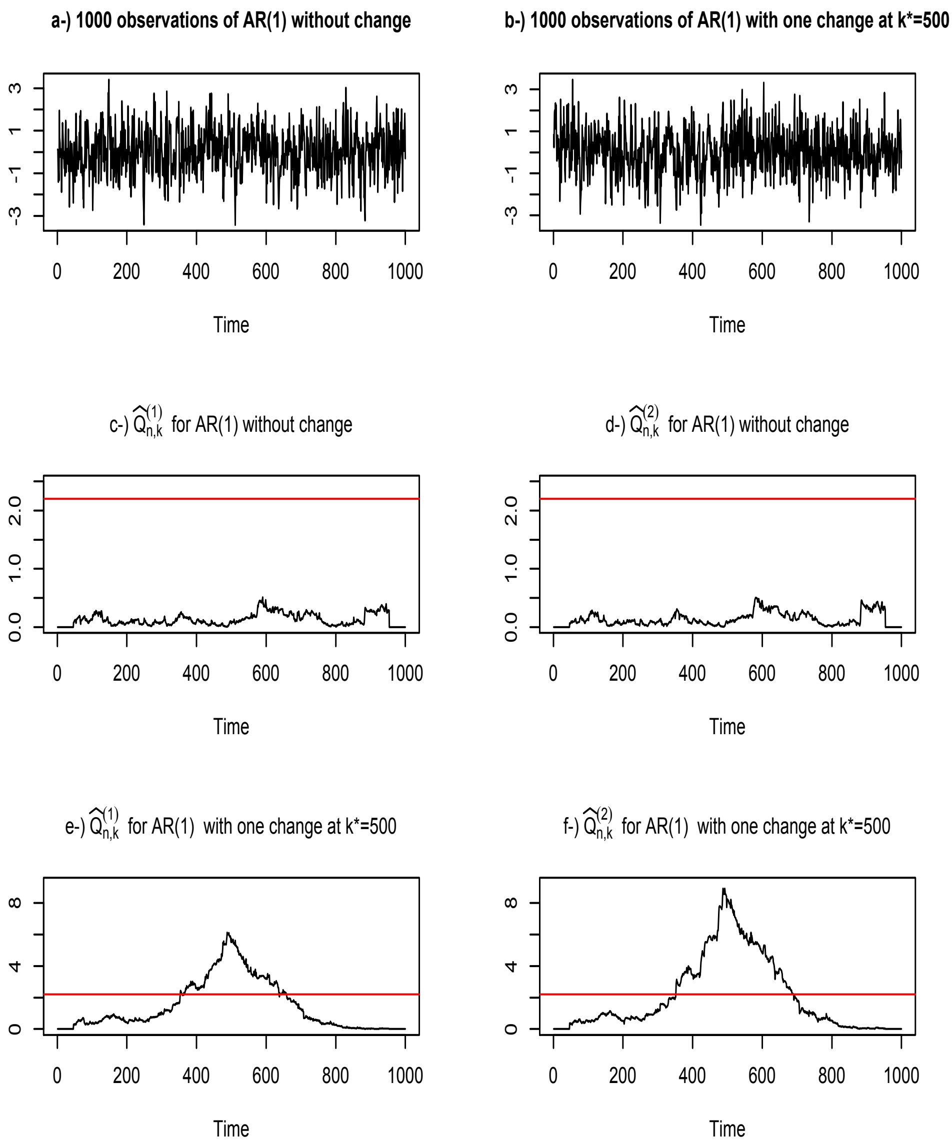

FIGURE 3.1 – Typical realization of 1000 observations of two AR(1) processes and the corresponding statistics $\widehat{Q}^{(1)}_{n,k}$ and $\widehat{Q}^{(2)}_{n,k}$ with $n = 1000$ and $v_n = [(\log n)^2]$. a-) is an AR(1) process without change, where the parameter $\phi_1 = 0.4$ is constant. b-) is an AR(1) process with one change at $k^* = 500$; the parameter $\phi_1 = 0.4$ changing to 0.2. c-), d-), e-) and f-) are their corresponding statistics $\widehat{Q}^{(1)}_{n,k}$ and $\widehat{Q}^{(2)}_{n,k}$.

$t^*_{K-1} < t^*_K = n$.
For establishing the consistency under the alternative, we add the following assumptions :

Assumption B : *there exists* $\tau^*_1, \ldots, \tau^*_{K^*-1}$ *with* $0 < \tau^*_1 < \ldots < \tau^*_{K^*-1} < 1$ *such that for* $j = 1, \cdots, K^*$, $t^*_j = [n\tau^*_j]$ *(where* $[x]$ *is the integer part of* x*).*
Assumption C : $\theta^*_1 \neq \theta^*_K$.

As we mentioned in the introduction, the property of stationary is lost after the first change. This situation (which is not common in the literature) increases the difficulty to

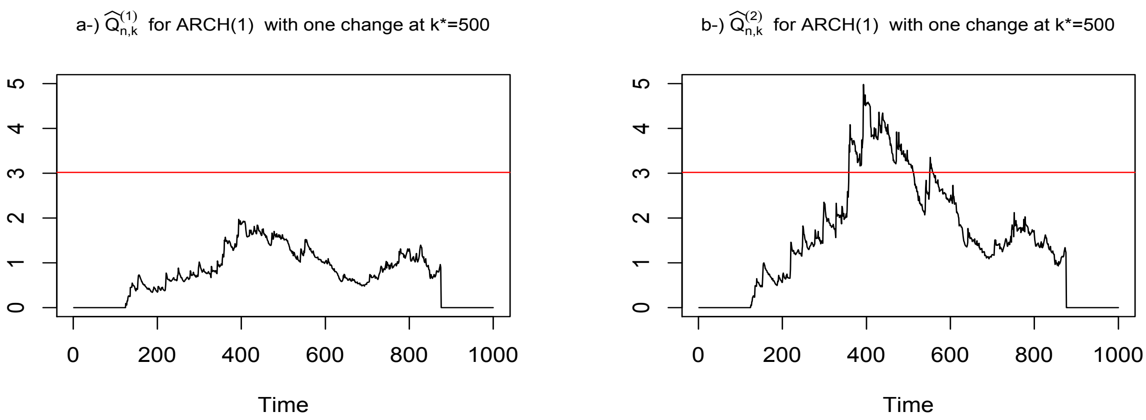

FIGURE 3.2 – Typical realization of the statistics $\widehat{Q}^{(1)}_{n,k}$ and $\widehat{Q}^{(2)}_{n,k}$ with $n = 1000$ and $v_n = [(\log n)^{5/2}]$ for an observations of ARCH(1) process. a-) and b-) are the case of ARCH(1) with the parameter $\theta_1 = (0.5, 0.7)$ changing to $(0.5, 0.4)$ at $k^* = 500$.

study asymptotic behavior under the alternative. Assumption B is classical when studying multiple change-point problem. In the case of one change in model ($K = 2$), Assumption C just means that $\theta^*_1 \neq \theta^*_2$, which is already holds because a change occurs. In the general situation, Assumption C is sufficient to prove the consistency in power. As we mentioned above, when the model depends on several parameters, the convergence of the estimator $\widehat{\theta}_n(T_n)$ is not ensured. So, we proceed as follows : since $\widehat{\theta}_n(T_{t^*_1})$ and $\widehat{\theta}_n(\overline{T}_{t^*_{K-1}})$ are the consistent estimators of θ^*_1 and θ^*_K respectively (see [8] and [9]), if $\theta^*_1 \neq \theta^*_K$ then, $\max(\|\widehat{\theta}_n(T_{t^*_1}) - \widehat{\theta}_n(T_n)\|, \|\widehat{\theta}_n(\overline{T}_{t^*_{K-1}}) - \widehat{\theta}_n(T_n)\|)$ will be asymptotically non-negative. This is enough to show that, the test statistic $\widehat{Q}_n$ diverges to infinity.

Theorem 3.3.2. Assume that $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$, $\mathbf{Var}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{K}(f_\theta, M_\theta, \Theta)$ are satisfied. Under the alternative H_1, if $\theta^*_1, \theta^*_K \in \overset{\circ}{\Theta}(4)$, then

$$\widehat{Q}_n \xrightarrow[n\to\infty]{a.s.} \infty.$$

Remark 3.3.2. 1-) Theorem 3.3.2 shows that the test is consistent in power.
2-) This procedure can also be used to detect multiple changes using iterated cumulative sums of squares (ICSS) type algorithm developed by Inclán and Tiao [37].

3.4 Some simulations results

In this section, we evaluate the performance of the procedure through empirical study. We compare our results with those obtained by Robbins *et al.* [63], Hušková *et al.* [35] (for the AR model) ; Lee and Na [54], Lee and Song [55] and Kulperger and Yu [48] (for the GARCH model). For asample size n, $\widehat{Q}_n$ is computed with $v_n = [(\ln n)^2]$ for AR model and $v_n = [(\ln n)^{5/2}]$ for GARCH model and is compared to the critical value of the test. In the following models, $(\xi_t)_{t\in\mathbb{Z}}$ are iid standard Gaussian random variables. In the financial time series, the distribution which has fatter tails than Gaussian distribution is often appropriate. So, in the GARCH model, the performances will also be evaluated using Student innovation ($t(8)$). The nominal level considered in the sequel is $\alpha = 0.05$.

3.4.1 Test for change in AR(p) models

Let us consider a AR(p) process : $X_t = \phi_0^* + \sum_{k=1}^{p} \phi_k^* X_{t-k} + \xi_t$ with $p \in \mathbb{N}^*$. The true parameter of the model is denoted by $\theta_0^* = (\phi_0^*, \phi_1^*, \cdots, \phi_p^*) \in \Theta$ where $\Theta = \{\theta = (\phi_0, \phi_1, \cdots, \phi_p) \in \mathbb{R}^{p+1} \ / \sum_{i=1}^{p} |\phi_i| < 1\}$. Since $M_\theta \equiv 1$, $\Theta(r) = \Theta$ for any $r \geq 1$. Assume $(X_1, \cdots, X_n)$ is observed, we have for any $\theta \in \Theta$,
$\widehat{q}_t(\theta) = \big(X_t - \phi_0 - \sum_{k=1}^{p} \phi_k X_{t-k}\big)^2$, $\dfrac{\partial \widehat{q}_t(\theta)}{\partial \theta} = -2\big(X_t - \phi_0 - \sum_{k=1}^{p} \phi_k X_{t-k}\big) \cdot (1, X_{t-1}, X_{t-2}, \cdots, X_{t-p})$.
Moreover, for $j = 1, \cdots, n$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial \phi_0 \partial \phi_j} = 2X_{t-j}$ and for $1 \leq i, j \leq n$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial \phi_i \partial \phi_j} = 2X_{t-i}X_{t-j}$.

Mean shift testing in AR(p)

Robbins *et al.* [63] consider the following problem of shift in mean

$$\begin{cases} X_t = \mu_0 + \epsilon_t & \text{for} \quad 1 \leq t \leq k^* \\ X_t = \mu_1 + \epsilon_t & \text{for} \quad t > k^* \end{cases}$$

where (ϵ_t) is a zero mean stationary series. Assume that (ϵ_t) is an AR(p) with mean zero i.e. $\epsilon_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \xi_t$.
Robbins *et al.* compare via simulation study, the performance of CUSUM, likelihood ratio and F_{max} test (see [63] for more details). An adjusted CUSUM test (which more powerful than the classical CUSUM test) is also proposed. We compare our procedure to the adjusted CUSUM test based on residuals. Let $(\widehat{\xi}_t)$ be the residuals of the models

$$X_t = \mu + \sum_{k=1}^{p} \phi_k X_{t-k} + \xi_t$$

and $\widehat{\sigma}_n^2$ a suitable estimator of the variance of the innovation. For $k = 1, \cdots, n-1$, denote

$$CUSUM_Z(k) = \frac{1}{\sqrt{n}} \Big(\sum_{t=1}^{k} \widehat{\xi}_k - \frac{k}{n} \sum_{t=1}^{n} \widehat{\xi}_k \Big) \quad \text{and} \quad \lambda_Z(k) = \frac{CUSUM_Z^2(k)}{\frac{k}{n}(1 - \frac{k}{n})}.$$

The residuals adjusted CUSUM test is based on the statistic

$$\frac{1}{\widehat{\sigma}_n^2} \max_{n\ell \leq k \leq nh} \lambda_Z(k)$$

where $0 < \ell < h < 1$. We will take $h = 1 - \ell$ with $\ell = 0.05$.
The tests are evaluated through a AR(1) process. We consider two cases : $\phi_1 = 0.6$ and $\phi_1 = -0.3$. For $n = 500, 1000, 1500$; we generate a sample $(X_1, \cdots, X_n)$ in the following situations : (i) no change in mean $\mu_0 = \mu_1 = 0$ and (ii) the mean changes from μ_0 to μ_1 at $n/2$. The Table 3.1 indicates the empirical levels and the empirical powers based of 200 replications for the tests using the statistics $\widehat{Q}_n$ and λ_Z.

		Procedure	$n = 500$	$n = 1000$	$n = 1500$
Emp. levels :	$\mu_0 = \mu_1 = 0\,;\ \phi_1 = 0.6$	$\widehat{Q}_n$ statistic	0.075	0.055	0.030
		λ_Z statistic	0.035	0.030	0.025
	$\mu_0 = \mu_1 = 0\,;\ \phi_1 = -0.3$	$\widehat{Q}_n$ statistic	0.040	0.035	0.035
		λ_Z statistic	0.055	0.040	0.035
Emp. powers :	$\mu_0 = 0;\ \mu_1 = 0.5\,;\ \phi_1 = 0.6$	$\widehat{Q}_n$ statistic	0.455	0.650	0.920
		λ_Z statistic	0.330	0.575	0.915
	$\mu_0 = 0;\ \mu_1 = 0.2\,;\ \phi_1 = -0.3$	$\widehat{Q}_n$ statistic	0.475	0.865	0.990
		λ_Z statistic	0.545	0.895	0.995

TABLE 3.1 – Empirical levels and powers at nominal level 0.05 of mean shift testing in AR(1) model in the cases $\phi_1 = 0.6$ and $\phi_1 = -0.3$.

Testing for change in the parameter of AR(p)

Consider the model a AR(p) process with mean zero

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + \xi_t, \quad \text{and denote} \quad \theta = (\phi_1, \cdots, \phi_p). \tag{3.3}$$

Hušková *et al.* [35] consider the problem of testing for parameter change in model (3.3). Their procedure is based on the sums of weighted residuals. With the above notation, denote for $t = p+1, \cdots, n \quad Y_t = (X_{t-1}, \cdots, X_{t-p})'$ and for $k = p+1, \cdots, n, \quad S_k = \sum_{t=p+1}^{k} Y_t \widehat{\xi}_t, \quad C_k = \sum_{t=p+1}^{k} Y_t Y_t'$ and $C_k^0 = C_n - C_k$. They discussed about three tests based on the statistics

$$W_n = \max_{p \le k \le n} (S_k' C_k^{-1} C_n (C_k^0)^{-1} S_k) / \widehat{\sigma}_n^2$$
$$W_n(\epsilon) = \max_{n\epsilon \le k \le n(1-\epsilon)} (S_k' C_k^{-1} C_n (C_k^0)^{-1} S_k) / \widehat{\sigma}_n^2$$
$$W_n(q) = \max_{p \le k \le n} \Big(\frac{S_k' C_n^{-1} S_k}{q(\frac{k}{k})} \Big) \Big/ \widehat{\sigma}_n^2$$

where $\epsilon \in (0, \frac{1}{2})$ and q is a positive weight function ; we will take $q = q_0 \equiv 1$. One can see in [36] that the statistic $W_n(q)$ performs better than W_n and $W_n(\epsilon)$; so, we will compare our procedure based on $\widehat{Q}_n$ to $W_n(q)$. The comparisons are based on AR(1) and AR(2) models, the parameter of the model is $\theta = \phi_1$ for AR(1) and $\theta = (\phi_1, \phi_2)$ for AR(2) process. Under alternative, the parameter θ_0 changes to θ_1 at $n/2$. Table 3.2 shows the empirical levels and the empirical powers based of 200 replications for the tests using the statistics $\widehat{Q}_n$ and $W_n(q)$.

As seen in Table 3.1, the empirical levels and powers of our procedure are asymptotically ($n \ge 1000$) close to those produced by adjusted CUSUM test for mean shift testing. For change in parameter, Table 3.2 shows that the two procedures using $\widehat{Q}_n$ and $W_n(q)$ statistics produce good empirical levels which approaching the nominal ones as n increases. One can see that, the empirical powers of our test are a little better than those obtained with sums of weighted residuals.

		Procedure	$n = 500$	$n = 1000$	$n = 1500$
Emp. levels :	$\theta_0 = 0.4$	$\widehat{Q}_n$ statistic	0.065	0.035	0.045
		$W_n(q)$ statistic	0.060	0.025	0.040
	$\theta_0 = -0.6$	$\widehat{Q}_n$ statistic	0.075	0.060	0.035
		$W_n(q)$ statistic	0.040	0.035	0.030
	$\theta_0 = (-0.3, 0.3)$	$\widehat{Q}_n$ statistic	0.125	0.040	0.060
		$W_n(q)$ statistic	0.105	0.035	0.060
Emp. powers :	$\theta_0 = 0.4;\ \theta_1 = 0.2$	$\widehat{Q}_n$ statistic	0.485	0.880	0.975
		$W_n(q)$ statistic	0.505	0.925	0.965
	$\theta_0 = -0.6;\ \theta_1 = -0.75$	$\widehat{Q}_n$ statistic	0.610	0.820	0.965
		$W_n(q)$ statistic	0.465	0.775	0.945
	$\theta_0 = (-0.3, 0.3)$ $\theta_1 = (-0.4, 0.4)$	$\widehat{Q}_n$ statistic	0.715	0.900	0.975
		$W_n(q)$ statistic	0.490	0.815	0.920

TABLE 3.2 – Empirical levels and powers at nominal level 0.05 of test for parameter change in AR(1) and AR(2) models.

3.4.2 Test for parameter change in GARCH(1,1) models

Consider the GARCH(1,1) model defined by :

$$\forall t \in \mathbb{Z}, \quad X_t = \sigma_t \xi_t \quad \text{with} \quad \sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$$

with $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*) \in \Theta \subset]0,\infty[\times[0,\infty[^2$ and satisfying $\alpha_1^* + \beta_1^* < 1$. The ARCH($\infty$) representation is $\sigma_t^2 = \alpha_0^*/(1-\beta_1^*) + \alpha_1^* \sum\limits_{k\geq 1} (\beta_1^*)^{k-1} X_{t-k}^2$.

For any $\theta \in \Theta$ and $t = 2, \cdots, n$, we have

$$\widehat{h}_\theta^t = \alpha_0/(1-\beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \sum_{k=2}^{t} \beta_1^{k-1} X_{t-k}^2 \quad \text{and} \quad \widehat{q}_t(\theta) = X_t^2 /\ \widehat{h}_\theta^t + \log(\widehat{h}_\theta^t).$$

Therefore, it follows that $\dfrac{\partial \widehat{q}_t(\theta)}{\partial \theta} = \dfrac{1}{\widehat{h}_\theta^t}\Big(1 - \dfrac{X_t^2}{\widehat{h}_\theta^t}\Big)\Big(\dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_0}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_1}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \beta_1}\Big)$ with $\partial \widehat{h}_\theta^t/\partial \alpha_1 = X_{t-1}^2 + \sum\limits_{k=2}^{t} \beta_1^{k-1} X_{t-k}^2$ $\partial \widehat{h}_\theta^t/\partial \alpha_0 = 1/(1-\beta_1)$, and $\partial \widehat{h}_\theta^t/\partial \beta_1 = \alpha_0/(1-\beta_1)^2 + \alpha_1 X_{t-2}^2 + \alpha_1 \sum\limits_{k=3}^{t} (k-1)\beta_1^{k-2} X_{t-k}^2$.
Let $\theta = (\alpha_0, \alpha_1, \beta_1) = (\theta_1, \theta_2, \theta_3) \in \Theta$, for $1 \leq i, j \leq 3$, we have

$$\frac{\partial^2 \widehat{q}_t(\theta)}{\partial \theta_i \partial \theta_j} = \frac{1}{(\widehat{h}_\theta^t)^2}\Big(\frac{2X_t^2}{\widehat{h}_\theta^t} - 1\Big)\frac{\partial \widehat{h}_\theta^t}{\partial \theta_i}\frac{\partial \widehat{h}_\theta^t}{\partial \theta_j} + \frac{1}{\widehat{h}_\theta^t}\Big(1 - \frac{X_t^2}{\widehat{h}_\theta^t}\Big)\frac{\partial^2 \widehat{h}_\theta^t}{\partial \theta_i \partial \theta_j}$$

with $\partial^2 \widehat{h}_\theta^t/\partial \alpha_0^2 = 0$, $\partial^2 \widehat{h}_\theta^t/\partial \alpha_0 \partial \alpha_1 = 0$, $\partial^2 \widehat{h}_\theta^t/\partial \alpha_1^2 = 0$, $\partial^2 \widehat{h}_\theta^t/\partial \alpha_1 \partial \beta_1 = X_{t-2}^2 + \sum\limits_{k=3}^{t} (k-1)\beta_1^{k-2} X_{t-k}^2$, $\partial^2 \widehat{h}_\theta^t/\partial \alpha_0 \partial \beta_1 = 1/(1-\beta_1)^2$ and $\partial \widehat{h}_\theta^t/\partial \beta_1^2 = 2\alpha_0/(1-\beta_1)^3 + 2\alpha_1 X_{t-3}^2 +$

$\alpha_1 \sum_{k=4}^{t} (k-1)(k-2)\beta_1^{k-3} X_{t-k}^2$.

Case of ARCH(1)

Assume $\beta_1 = 0$ and $\theta = (\alpha_0, \alpha_1)$. As we mentioned in the introduction, Lee and Na [54] proposed a test based on the conditional least-squares estimator. For $k = 1, \cdots, n-1$, the estimator $\widehat{\theta}_k = (\widehat{\alpha}_{0,k}, \widehat{\alpha}_{1,k})$ computed on the segment $\{1, \cdots, k\}$ is obtained by minimizing the conditional sum of squares

$$\sum_{t=1}^{k} (X_t - \alpha_0 - \alpha_1 X_{t-1})^2.$$

Denote $\text{ł}_t(\theta) = \frac{1}{2}(X_t - \alpha_0 - \alpha_1 X_{t-1})^2$. Define the matrix

$$\widehat{V} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 \text{ł}_t(\widehat{\theta}_n)}{\partial\theta\partial\theta'} \quad \text{and} \quad \widehat{W} = \frac{1}{n} \sum_{t=1}^{n} \Big(\frac{\partial \text{ł}_t(\widehat{\theta}_n)}{\partial\theta}\Big)\Big(\frac{\partial \text{ł}_t(\widehat{\theta}_n)}{\partial\theta}\Big)'.$$

Under H_0 and some regularity conditions, $\widehat{V}^{-1}\widehat{W}\widehat{V}^{-1}$ is a consistent estimator of the asymptotic covariance of $\widehat{\theta}_n$ (see [43]). The test of Lee and Na is based on the statistic

$$LS_n = \max_{2 \le k \le n} \Big(\frac{k^2}{n}(\widehat{\theta}_k - \widehat{\theta}_n)\widehat{V}\widehat{W}^{-1}\widehat{V}(\widehat{\theta}_k - \widehat{\theta}_n)\Big).$$

For $n = 500, 1000, 1500$; we generate a sample $(X_1, \cdots, X_n)$ of ARCH(1) in the following situations : (i) there is no change, the parameter of the model θ_0 is constant and (ii) there is one change, θ_0 changes to θ_1 at $n/2$. Table 3.3 indicates the empirical levels and the empirical powers based of 200 replications for the tests using the statistics $\widehat{Q}_n$, $\widehat{Q}_n^{(0)}$ (see (3.2), proposed by Lee and Song [55]), LS_n and residuals CUSUM statistic using residual CUSUM test statistic(see Kulperger and Yu [48]).

Case of GARCH(1,1)

Now, $\theta = (\alpha_0, \alpha_1, \beta_1)$. The performance of the procedure using the statistics $\widehat{Q}_n$, $\widehat{Q}_n^{(0)}$ and the one using residual CUSUM test statistic (see Kulperger and Yu [48]) are evaluated. Table 3.4 indicates the empirical levels and powers of these procedure based on 100 replications; the empirical powers are computed when the parameter θ_0 changes to θ_1 at $n/2$.

As shown in Table 3.3 and 3.4, apart from the procedure based on conditional least-squares estimator (in ARCH(1)), the nominal levels of these procedure are close to the nominal ones when $n = 1500$. The residual CUSUM test produces less size distortion than others procedure. As pointed out by many authors (for instance Lee and Na [54]), the residual CUSUM test in GARCH models it intrinsically detects change of unconditional variance. So, see the situation of ARCH(1) where $\theta_0 = (0.3, 0.1)$ and $\theta_1 = (0.2, 0.4)$. A change occurs in the parameter but the unconditional variance (of the stationary model) remains the same. One can see that the residual CUSUM test produced very poor powers in this situation. It is the case for GARCH(1,1) in the situation where $\theta_0 = (1, 0.4, 0.1)$ and $\theta_1 = (1, 0.2, 0.3)$. In ARCH model, our test outperforms the procedure based on conditional least-squares estimator. This procedure produces large size distortion and very

		Procedure	$n = 500$	$n = 1000$	$n = 1500$
Emp. levels :	$\theta_0 = (0.3, 0.1)$	$\widehat{Q}_n$ statistic	0.075	0.050	0.045
		$\widehat{Q}_n^{(0)}$ statistic	0.055	0.065	0.060
		LS_n statistic	0.100	0.095	0.095
		CUSUM test	0.045	0.055	0.050
	$\theta_0 = (0.5, 0.7)$	$\widehat{Q}_n$ statistic	0.080	0.075	0.055
		$\widehat{Q}_n^{(0)}$ statistic	0.070	0.070	0.060
		LS_n statistic	0.150	0.100	0.090
		CUSUM test	0.015	0.040	0.045
Emp. powers :	$\theta_0 = (0.3, 0.1)$ $\theta_1 = (0.2, 0.1)$	$\widehat{Q}_n$ statistic	0.665	0.960	0.990
		$\widehat{Q}_n^{(0)}$ statistic	0.655	0.960	0.990
		LS_n statistic	0.580	0.935	0.970
		CUSUM test	0.675	0.970	0.980
	$\theta_0 = (0.3, 0.1)$ $\theta_1 = (0.2, 0.4)$	$\widehat{Q}_n$ statistic	0.575	0.860	0.970
		$\widehat{Q}_n^{(0)}$ statistic	0.455	0.855	0.950
		LS_n statistic	0.465	0.695	0.725
		CUSUM test	0.120	0.180	0.260
	$\theta_0 = (0.5, 0.7)$ $\theta_1 = (0.5, 0.4)$	$\widehat{Q}_n$ statistic	0.415	0.690	0.765
		$\widehat{Q}_n^{(0)}$ statistic	0.180	0.445	0.620
		LS_n statistic	0.150	0.170	0.155
		CUSUM test	0.200	0.405	0.490
	$\theta_0 = (0.5, 0.7)$ $\theta_1 = (0.75, 0.55)$	$\widehat{Q}_n$ statistic	0.465	0.560	0.830
		$\widehat{Q}_n^{(0)}$ statistic	0.280	0.500	0.785
		LS_n statistic	0.215	0.235	0.305
		CUSUM test	0.250	0.375	0.560

Table 3.3 – Empirical levels and powers at nominal level 0.05 of test for parameter change in ARCH(1) model.

		Procedure	$n = 500$	$n = 1000$	$n = 1500$
Emp. levels :	$\theta_0 = (1, 0.4, 0.1)$	$\widehat{Q}_n$ statistic	0.09 (0.19)	0.08 (0.15)	0.04 (0.09)
		$\widehat{Q}_n^{(0)}$ statistic	0.07 (0.18)	0.06 (0.16)	0.05 (0.15)
		CUSUM test	0.04 (0.05)	0.04 (0.03)	0.05 (0.02)
Emp. powers :	$\theta_1 = (0.5, 0.4, 0.1)$	$\widehat{Q}_n$ statistic	0.60 (0.53)	0.92 (0.85)	0.98 (0.93)
		$\widehat{Q}_n^{(0)}$ statistic	0.81 (0.72)	0.96 (0.95)	0.97 (0.97)
		CUSUM test	0.80 (0.61)	0.97 (0.93)	0.99 (0.96)
	$\theta_1 = (1, 0.2, 0.3)$	$\widehat{Q}_n$ statistic	0.25 (0.37)	0.53 (0.65)	0.57 (0.68)
		$\widehat{Q}_n^{(0)}$ statistic	0.12 (0.29)	0.35 (0.61)	0.60 (0.71)
		CUSUM test	0.05 (0.02)	0.19 (0.11)	0.22 (0.18)

Table 3.4 – Empirical levels and powers at nominal level 0.05 of test for parameter change in GARCH(1,1) model. Figures in brackets are the results obtained when the innovation follows a Student distribution of 8 degrees of freedom.

poor powers in some case. Comparing to the test based on $\widehat{Q}_n^{(0)}$, one can see that the results obtained with our procedure are more accurate in many case. As seen in Table 3.4, the test based on $\widehat{Q}_n$ and $\widehat{Q}_n^{(0)}$ produce large size distortion when the innovation follows Student distribution. But, we can see that the empirical performance of these procedure increases as n increases.

3.4.3 Test for parameter change in TARCH(1) models

Consider a TARCH(1) process :

$$X_t = \sigma_t \xi_t \ , \ \sigma_t = b_0 + b_1^+ \max(X_{t-1}, 0) - b_1^- \min(X_{t-1}, 0), \ \forall t \in \mathbb{Z}.$$

The vector of parameter is $\theta = (b_0, b_1^+, b_1^-) \in (0, \infty) \times [0, \infty)^2$. For $n = 500, 1000, 1500$; we generate a trajectory $(X_1, \cdots, X_n)$ of TARCH(1) which the parameter θ_0 remains constant under null hypothesis and changes to θ_1 at $n/2$ under alternative. The Table 3.5 shows the empirical levels and powers based on 100 replications of the procedure using statistic $\widehat{Q}_n$.

		$n = 500$	$n = 1000$	$n = 1500$
Emp. levels :	$\theta_0 = (0.5, 0.01, 0.03)$	0.18	0.14	0.11
Emp. powers :	$\theta_0 = (0.5, 0.01, 0.03)$; $\theta_1 = (0.3, 0.01, 0.03)$	0.58	0.77	0.86
	$\theta_0 = (0.5, 0.01, 0.03)$; $\theta_1 = (0.5, 0.07, 0.03)$	0.53	0.81	0.84

Table 3.5 – Empirical levels and powers at nominal level 0.05 of test for parameter change in TARCH(1) model.

Notice that, testing for parameter change in threshold model is a real challenge in statistic. The difficulty increases when there exists asymmetric effect in the model. Any procedure will be able to not confuse the asymmetric effect and a change in model. We consider an example with moderate asymmetric effect ; the Table 3.5 shows that our procedure provides more acceptable results (when n increases) in this case.

3.5 Conclusion

This paper is devoted to the problem of testing for parameter change on a large class of causals models. We construct a statistic which takes into account the change-point alternative. It is compared to some existing procedures through a simulation study. The results show that our procedure works well for mean shift testing in AR(p) model as adjusted CUSUM test. For changes in parameter, our procedure is more powerful than the procedure based on sums of weighted residuals proposed by Hušková *et al.* [35]. For (G)ARCH model, it is shown that the residual CUSUM test outperforms only when the change in parameter induces a change in unconditional variance, in other case, it produces very poor power. It is also shown that our procedure is more powerful than the one based on conditional least-squares estimator in ARCH(1). The comparison with the procedure based on $\widehat{Q}_n^{(0)}$ (proposed by Lee and Song [55]) is in favor of our test in many cases. According to all the results of the simulation, we recommend to use our procedure when $n \geq 1000$.

3.6 Proofs of the main results

Let $(\psi_n)_n$ and $(r_n)_n$ be sequences of random variables. Throughout this section, we use the notation $\psi_n = o_P(r_n)$ to mean : for all $\varepsilon > 0$, $P(|\psi_n| \geq \varepsilon |r_n|) \to 0$ as $n \to \infty$. Write $\psi_n = O_P(r_n)$ to mean : for all $\varepsilon > 0$, there exists $C > 0$ such that $P(|\psi_n| \geq C|r_n|) \leq \varepsilon$ for n large enough.

3.6.1 Some preliminary results

First, let us prove useful technical lemmas.

Under the null hypothesis H_0 the observations $(X_1, \cdots, X_n)$ belong to the class $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0}, f_{\theta_0})$, define the matrix $G := \mathbb{E}\Big[\dfrac{\partial q_0(\theta_0)}{\partial \theta}\dfrac{\partial q_0(\theta_0)}{\partial \theta}'\Big]$ (where $'$ denotes the transpose) and $F := \mathbb{E}\Big[\dfrac{\partial^2 q_0(\theta_0)}{\partial \theta \partial \theta'}\Big]$. Under assumption **Var**, F is a non-singular matrix (see [8]).

Lemma 3.6.1. Assume the functions $\theta \mapsto M_\theta$ and $\theta \mapsto f_\theta$ are 2-times continuously differentiable on Θ. Under the null hypothesis $\mathbf{D}(\Theta)$ and **Var**, G is a symmetric, positive definite matrix.

Proof. It is clear that G is symmetric. Moreover, for $1 \leq i \leq d$, we have :
$\dfrac{\partial q_0(\theta_0)}{\partial \theta_i} = -2\dfrac{\xi_0}{\sqrt{h^0_{\theta_0}}}\dfrac{\partial f^0_{\theta_0}}{\partial \theta_i} - \dfrac{\xi_0^2}{h^0_{\theta_0}}\dfrac{\partial h^0_{\theta_0}}{\partial \theta_i} + \dfrac{1}{h^0_{\theta_0}}\dfrac{\partial h^0_{\theta_0}}{\partial \theta_i}$. Thus, using independence of ξ_0 and $X_{-1},\ X_{-2}, \cdots$ we obtain

$$\mathbb{E}\Big[\frac{\partial q_0(\theta_0)}{\partial \theta}'\frac{\partial q_0(\theta_0)}{\partial \theta}\Big] = 4\mathbb{E}\Big[\frac{1}{h^0_{\theta_0}}\frac{\partial f^0_{\theta_0}}{\partial \theta}'\frac{\partial f^0_{\theta_0}}{\partial \theta}\Big] + \mathbb{E}\Big((\xi_0^2-1)^2\Big)\mathbb{E}\Big[\frac{1}{(h^0_{\theta_0})^2}\frac{\partial h^0_{\theta_0}}{\partial \theta}'\frac{\partial h^0_{\theta_0}}{\partial \theta}\Big]. \quad (3.4)$$

Since $\mathbb{E}\xi_0^2 = 1$, it is easy to see that $\mathbb{E}\Big((\xi_0^2-1)^2\Big) > 0$. Under **Var**, one of the two matrix of the right-hand side of relation (3.4) is positive definite and the other is positive semi-definite. Thus, G is positive definite. □

Now, recall that $F := I\!E\Big[\dfrac{\partial^2 q_0(\theta_0)}{\partial\theta\partial\theta'}\Big]$. Let $T \subset \{1, \cdots, n\}$. For any $\theta \in \Theta$ and $i = 1, \cdots, d$, by Taylor expansion of $\partial L_n(T, \theta_0)/\partial\theta_i$, there exist $\overline{\theta}_{n,i} \in [\theta_0, \theta]$ such that :

$$\frac{\partial L_n(T,\theta)}{\partial\theta_i} = \frac{\partial L_n(T,\theta_0)}{\partial\theta_i} + \frac{\partial^2 L_n(T,\overline{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\theta-\theta_0) \tag{3.5}$$

where $[a,b] = \{\lambda a + (1-\lambda)b \,;\, \lambda \in [0,1]\}$. Denote $\overline{F}_n(T,\theta) = -2\big(\dfrac{1}{card(T)}\dfrac{\partial^2 L_n(T,\overline{\theta}_{n,i})}{\partial\theta\partial\theta_i}\big)_{1\leq i\leq d}$. Then, (3.5) implies,

$$Card(T)\overline{F}_n(T,\theta)(\theta-\theta_0) = -2\Big(\frac{\partial L_n(T,\theta)}{\partial\theta} - \frac{\partial L_n(T,\theta_0)}{\partial\theta}\Big). \tag{3.6}$$

Similarly, for any $\theta \in \Theta$ we can find a matrix $\widetilde{F}_n(T,\theta)$ such that

$$Card(T)\widetilde{F}_n(T,\theta)(\theta-\theta_0) = -2\Big(\frac{\partial \widehat{L}_n(T,\theta)}{\partial\theta} - \frac{\partial \widehat{L}_n(T,\theta_0)}{\partial\theta}\Big). \tag{3.7}$$

With $\theta = \widehat{\theta}_n(T)$ in (3.7) and using the fact that $\partial\widehat{L}_n(T,\widehat{\theta}_n(T))/\partial\theta = 0$ (because $\widehat{\theta}_n(T)$ is a local extremum of $\widehat{L}_n(T,\cdot)$), it comes

$$Card(T)\widetilde{F}_n(T,\widehat{\theta}_n(T))(\widehat{\theta}_n(T)-\theta_0) = 2\frac{\partial\widehat{L}_n(T,\theta_0)}{\partial\theta}. \tag{3.8}$$

Remark 3.6.1. If $Card(T) \underset{n\to\infty}{\longrightarrow} \infty$ and $\theta = \theta(n) \underset{n\to\infty}{\longrightarrow} \theta_0$, then $\overline{F}_n(T,\theta) \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} F$ and $\widetilde{F}_n(T,\theta) \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} F$ (see [8] and [9]). In particular, if $Card(T) \underset{n\to\infty}{\longrightarrow} \infty$, then $\overline{F}_n(T,\widehat{\theta}_n(T)) \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} F$ and $\widetilde{F}_n(T,\widehat{\theta}_n(T)) \overset{a.s.}{\underset{n\to\infty}{\longrightarrow}} F$.

Lemma 3.6.2. Under assumptions of Theorem 3.3.1

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\big\|k\big(\widetilde{F}_n(T_k,\widehat{\theta}_n(T_k)) - F\big)\big(\widehat{\theta}_n(T_k)-\theta_0\big)\big\| = o_P(1).$$

Proof. For $k \in \Pi_n$, we know that $\sqrt{k}(\widehat{\theta}_n(T_k)) - \theta_0)$ converges in distribution to the Gaussian law as $n \longrightarrow \infty$ (see Theorem 2 of [8]). Therefore, $\max\limits_{k\in\Pi_n}\big\|\sqrt{k}(\widehat{\theta}_n(T_k)-\theta_0)\big\| = O_P(1)$. Remark 3.6.1 implies that $\max\limits_{k\in\Pi_n}\big\|\widetilde{F}_n(T_k,\widehat{\theta}_n(T_k)) - F\big\| = o(1)$ a.s. Thus

$$\begin{aligned}\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\big\|k\big(\widetilde{F}_n(T_k,\widehat{\theta}_n(T_k)) - F\big)\big(\widehat{\theta}_n(T_k)-\theta_0\big)\big\| &\leq \max_{k\in\Pi_n}\big\|\widetilde{F}_n(T_k,\widehat{\theta}_n(T_k)) - F\big\| \\ &\quad\times \max_{k\in\Pi_n}\big\|\sqrt{k}(\widehat{\theta}_n(T_k)-\theta_0)\big\| \qquad (3.9)\\ &= o(1)O_P(1) \text{ a.s.} \\ &= o_P(1). \ \square\end{aligned}$$

Under assumptions of Theorem 3.3.1, the matrix G is invertible. Denote $\Sigma = FG^{-1}F$

$$Q_n^{(1)} := \max_{k\in\Pi_n} Q_{n,k}^{(1)} \quad \text{where} \quad Q_{n,k}^{(1)} := \frac{k^2}{n}\big(\widehat{\theta}_n(T_k)-\widehat{\theta}_n(T_n)\big)'\Sigma\big(\widehat{\theta}_n(T_k)-\widehat{\theta}_n(T_n)\big) \quad \text{and}$$

$$Q_n^{(2)} := \max_{k\in\Pi_n} Q_{n,k}^{(2)} \quad \text{where} \quad Q_{n,k}^{(2)} := \frac{(n-k)^2}{n}\big(\widehat{\theta}_n(\overline{T}_k)-\widehat{\theta}_n(T_n)\big)'\Sigma\big(\widehat{\theta}_n(\overline{T}_k)-\widehat{\theta}_n(T_n)\big).$$

Lemma 3.6.3. Under assumptions of Theorem 3.3.1

$$\max_{k\in\Pi_n} |\widehat{Q}^{(j)}_{n,k} - Q^{(j)}_{n,k}| = o_P(1) \text{ for } j = 1, 2.$$

Proof. The proof is provided for $j = 1$, proceed similarly for $j = 2$. For any $k \in \Pi_n$, we have

$$\begin{aligned}
|\widehat{Q}^{(1)}_{n,k} - Q^{(1)}_{n,k}| &\leq \frac{k^2}{n} \|\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)\|^2 \|\widehat{\Sigma}_{n,k} - \Sigma\| \\
&\leq 2\frac{k^2}{n} (\|\widehat{\theta}_n(T_k) - \theta_0\|^2 + \|\widehat{\theta}_n(T_n) - \theta_0\|^2) \|\widehat{\Sigma}_{n,k} - \Sigma\| \\
&\leq 2(\|\sqrt{k}(\widehat{\theta}_n(T_k) - \theta_0)\|^2 + \|\sqrt{n}(\widehat{\theta}_n(T_n) - \theta_0)\|^2) \|\widehat{\Sigma}_{n,k} - \Sigma\|. \qquad (3.10)
\end{aligned}$$

Since $k \in \Pi_n$, $k, n - k \longrightarrow \infty$ as $n \longrightarrow \infty$. Therefore, $\sqrt{k}(\widehat{\theta}_n(T_k) - \theta_0) = O_P(1)$ as $n \longrightarrow \infty$, $\sqrt{n}(\widehat{\theta}_n(T_n) - \theta_0) = O_P(1)$, $\widehat{F}_n(T_k) \xrightarrow[n\to\infty]{a.s.} F$, $\widehat{F}_n(\overline{T}_k) \xrightarrow[n\to\infty]{a.s.} F$, $\widehat{G}_n(T_k) \xrightarrow[n\to\infty]{a.s.} G$ and $\widehat{G}_n(\overline{T}_k) \xrightarrow[n\to\infty]{a.s.} G$ which is invertible. Thus, for n large enough, $\widehat{G}_n(T_k)$ and $\widehat{G}_n(\overline{T}_k)$ are invertible. It follows that as $n \longrightarrow \infty$,

$$\begin{aligned}
\|\widehat{\Sigma}_{n,k} - \Sigma\| &= \|\frac{k}{n} \widehat{F}_n(T_k)\widehat{G}_n(T_k)^{-1}\widehat{F}_n(T_k) + \frac{n-k}{n} \widehat{F}_n(\overline{T}_k)\widehat{G}_n(\overline{T}_k)^{-1}\widehat{F}_n(\overline{T}_k) - FG^{-1}F\| \\
&= \|\frac{k}{n}(\frac{k}{n} \widehat{F}_n(T_k)\widehat{G}_n(T_k)^{-1}\widehat{F}_n(T_k) - FG^{-1}F) \\
&\qquad\qquad + \frac{n-k}{n}(\widehat{F}_n(\overline{T}_k)\widehat{G}_n(\overline{T}_k)^{-1}\widehat{F}_n(\overline{T}_k) - FG^{-1}F)\| \\
&\leq \|\widehat{F}_n(T_k)\widehat{G}_n(T_k)^{-1}\widehat{F}_n(T_k) - FG^{-1}F\| + \|\widehat{F}_n(\overline{T}_k)\widehat{G}_n(\overline{T}_k)^{-1}\widehat{F}_n(\overline{T}_k) - FG^{-1}F\| \\
&= o(1) \text{ a.s.}
\end{aligned}$$

Therefore, (3.10) implies $\max_{k\in\Pi_n} |\widehat{Q}^{(1)}_{n,k} - Q^{(1)}_{n,k}| = o_P(1)$. □

Lemma 3.6.4. Under assumptions of Theorem 3.3.1

$$\frac{-2}{\sqrt{n}} \frac{\partial L_n(T_{[n\tau]}, \theta_0)}{\partial \theta} \xrightarrow{\mathcal{D}} W_G(\tau) \quad \text{in} \quad D([0,1], \mathbb{R}^d)$$

where W_G is a d-dimensional Gaussian process with zero mean and covariance matrix $\min(\tau, s)G$.

Proof. Recall that $-2\frac{\partial L_n(T_{[n\tau]}, \theta_0)}{\partial \theta} = \sum_{t=1}^{[n\tau]} \frac{\partial q_t(\theta_0)}{\partial \theta}$. Denote $\mathcal{F}_t = \sigma(X_{t-1}, \cdots)$. Since X is stationary and ergodic, it is the same for the process $(\frac{\partial q_t(\theta_0)}{\partial \theta})_{t\in\mathbb{Z}}$. Moreover, $(\frac{\partial q_t(\theta_0)}{\partial \theta}, \mathcal{F}_t)$ is a square integrable martingale difference process (see [8]) with covariance matrix G. Then, the result follow by using Theorem 23.1 Billingsley (1968) (see [18] page 206). □

Lemma 3.6.5. Under assumptions of Theorem 3.3.1

$$\frac{-2}{\sqrt{n}} G^{-1/2}\Big(\frac{\partial L_n(T_{[n\tau]}, \theta_0)}{\partial \theta} - \frac{[n\tau]}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta}\Big) \xrightarrow{\mathcal{D}} W_d(\tau) \quad \text{in} \quad D([0,1], \mathbb{R}^d)$$

where W_d is a d-dimensional Brownian bridge.

Proof. By Lemma 3.6.4, it comes

$$\frac{-2}{\sqrt{n}}\Big(\frac{\partial L_n(T_{[n\tau]},\theta_0)}{\partial\theta}-\frac{[n\tau]}{n}\frac{\partial L_n(T_n,\theta_0)}{\partial\theta}\Big)\xrightarrow{\mathcal{D}}\quad W_G(\tau)-\tau W_G(1)\quad\text{in}\ \ D([0,1],\mathbb{R}^d).$$

Since the covariance matrix of the process $\{W_G(\tau)-\tau W_G(1),\ 0\le\tau\le 1\}$ is $(\min(\tau,s)-\tau s)G$, the covariance matrix of the process $\{G^{-1/2}(W_G(\tau)-\tau W_G(1)),\ 0\le\tau\le 1\}$ is $(\min(\tau,s)-\tau s)I_d$ (where I_d is the d-dimensional identity matrix). Therefore, the process is equal (in distribution) to a d-dimensional Brownian bridge and the result follows. □

Lemma 3.6.6. Under assumptions of Theorem 3.3.1

$$\frac{-2}{\sqrt{n}}G^{-1/2}\frac{\partial\widehat{L}_n(T_{[n\tau]},\widehat{\theta}_n(T_n))}{\partial\theta}\xrightarrow{\mathcal{D}}\quad W_d(\tau)\quad\text{in}\ \ D([0,1],\mathbb{R}^d).$$

Proof. From [8], we have $\frac{1}{\sqrt{n}}\|\frac{\partial L_n(T_n,\cdot)}{\partial\theta}-\frac{\partial\widehat{L}_n(T_n,\cdot)}{\partial\theta}\|_\Theta=o_P(1)$. This implies,

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|\frac{\partial L_n(T_k,\cdot)}{\partial\theta}-\frac{\partial\widehat{L}_n(T_k,\cdot)}{\partial\theta}\|_\Theta=o_P(1).\qquad(3.11)$$

Let $k\in\Pi_n$. Applying (3.6) with $T=T_k$ and $\theta=\widehat{\theta}_n(T_n)$, we have

$$k\overline{F}_n(T_k,\widehat{\theta}_n(T_n))(\widehat{\theta}_n(T_n)-\theta_0)=-2\big(\frac{\partial L_n(T_k,\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{\partial L_n(T_k,\theta_0)}{\partial\theta}\big).$$

By plugging it in (3.11), we have

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|\frac{\partial\widehat{L}_n(T_k,\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{\partial L_n(T_k,\theta_0)}{\partial\theta}+\frac{1}{2}k\overline{F}_n(T_k,\widehat{\theta}_n(T_n))(\widehat{\theta}_n(T_n)-\theta_0)\|=o_P(1).\qquad(3.12)$$

But, by Remark 3.6.1, it comes that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|k(\overline{F}_n(T_k,\widehat{\theta}_n(T_n))-\overline{F}_n(T_n,\widehat{\theta}_n(T_n)))(\widehat{\theta}_n(T_n)-\theta_0)\|\\&\qquad\le\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|k(\overline{F}_n(T_k,\widehat{\theta}_n(T_n))-\overline{F}_n(T_n,\widehat{\theta}_n(T_n)))\|\times\|\sqrt{n}(\widehat{\theta}_n(T_n)-\theta_0)\|\\&\qquad=o(1)O_P(1)\quad\text{a.s.}\\&\qquad=o_P(1).\end{aligned}$$

Thus, (3.12) becomes

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|\frac{\partial\widehat{L}_n(T_k,\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{\partial L_n(T_k,\theta_0)}{\partial\theta}+\frac{1}{2}k\overline{F}_n(T_n,\widehat{\theta}_n(T_n))(\widehat{\theta}_n(T_n)-\theta_0)\|=o_P(1).\qquad(3.13)$$

Applying (3.6) with $T=T_n$, $\theta=\widehat{\theta}_n(T_n)$, and using $(1/\sqrt{n})(\partial L_n(T_n,\widehat{\theta}_n(T_n))/\partial\theta)=o_P(1)$ (see [8]), it follows

$$\overline{F}_n(T_n,\widehat{\theta}_n(T_n))(\widehat{\theta}_n(T_n)-\theta_0)=\frac{2}{n}\frac{\partial L_n(T_n,\theta_0)}{\partial\theta}+o_P(\frac{1}{\sqrt{n}}).\qquad(3.14)$$

Therefore, (3.13) becomes

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\|\frac{\partial\widehat{L}_n(T_k,\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{\partial L_n(T_k,\theta_0)}{\partial\theta}+\frac{k}{n}\frac{\partial L_n(T_n,\theta_0)}{\partial\theta}\|=o_P(1).\qquad(3.15)$$

Now, let $0 < \tau < 1$, for large value of n, we have $[\tau n] \in \Pi_n$; write

$$\frac{-2}{\sqrt{n}} G^{-1/2} \frac{\partial \widehat{L}_n(T_{[n\tau]}, \widehat{\theta}_n(T_n))}{\partial \theta} = \frac{-2}{\sqrt{n}} G^{-1/2} \Big[\frac{\partial \widehat{L}_n(T_{[n\tau]}, \widehat{\theta}_n(T_n))}{\partial \theta} - \Big(\frac{\partial L_n(T_{[n\tau]}, \theta_0)}{\partial \theta} - \frac{[n\tau]}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} \Big) + \Big(\frac{\partial L_n(T_{[n\tau]}, \theta_0)}{\partial \theta} - \frac{[n\tau]}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} \Big) \Big]$$

and the result follows by using (3.15) and Lemma 3.6.5. □

3.6.2 Proof of Theorem 3.3.1 and Theorem 3.3.2

Proof of Theorem 3.3.1 .
We give the proof for $j = 1$, proceed similarly for $j = 2$. By Lemma 3.6.3, Theorem 3.3.1 is established if $Q_n^{(1)} \xrightarrow[n \to \infty]{\mathcal{D}} \sup_{0 \leq \tau \leq 1} \|W_d(\tau)\|^2$. Using (3.11), (3.8) with $T = T_k$ and Lemma 3.6.2 it follows

$$\begin{aligned} \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{\partial L_n(T_k, \theta_0)}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \theta_0) \Big\| &= \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{\partial \widehat{L}_n(T_k, \theta_0)}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \theta_0) \Big\| + o_P(1) \\ &= \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{1}{2} k \widetilde{F}_n(T_k, \widehat{\theta}_n(T_k))(\widehat{\theta}_n(T_n) - \theta_0) - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \theta_0) \Big\| + o_P(1) \\ &= \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{1}{2} k (\widetilde{F}_n(T_k, \widehat{\theta}_n(T_k)) - F)(\widehat{\theta}_n(T_n) - \theta_0) \Big\| + o_P(1) = o_P(1). \end{aligned} \tag{3.16}$$

Using (3.15) and 3.16, we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{\partial L_n(T_k, \widehat{\theta}_n(T_n))}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)) \Big\| \\ &\quad = \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{\partial L_n(T_k, \theta_0)}{\partial \theta} - \frac{k}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)) \Big\| + o_P(1) \\ &\quad = \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \theta_0) - \frac{k}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)) \Big\| + o_P(1) \\ &\quad = \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{1}{2} kF\, (\widehat{\theta}_n(T_n) - \theta_0) - \frac{k}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} \Big\| + o_P(1) \\ &\quad \leq \sqrt{n} \Big\| \frac{1}{2} F\, (\widehat{\theta}_n(T_n) - \theta_0) - \frac{1}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} \Big\| + o_P(1). \end{aligned} \tag{3.17}$$

Note that

$$\begin{aligned} \big\| \sqrt{n}(F - \overline{F}_n(T_n, \widehat{\theta}_n(T_n)))\, (\widehat{\theta}_n(T_n) - \theta_0) \big\| &\leq \big\| F - \overline{F}_n(T_n, \widehat{\theta}_n(T_n)) \big\| \, \big\| \sqrt{n}(\widehat{\theta}_n(T_n) - \theta_0) \big\| \\ &= o(1) O_P(1) \quad \text{a.s.} \\ &= o_P(1). \end{aligned}$$

By plugging it in (3.17) and applying (3.6) with $T = T_n$ and $\theta = \widehat{\theta}_n(T_n)$, we have

$$\begin{aligned} \frac{1}{\sqrt{n}} \max_{k \in \Pi_n} \Big\| \frac{\partial L_n(T_k, \widehat{\theta}_n(T_n))}{\partial \theta} - \frac{1}{2} kF\, (\widehat{\theta}_n(T_k) - \widehat{\theta}_n(T_n)) \Big\| \leq \sqrt{n} \Big\| & \frac{1}{2} \overline{F}_n(T_n, \widehat{\theta}_n(T_n))(\widehat{\theta}_n(T_n) - \theta_0) \\ & - \frac{1}{n} \frac{\partial L_n(T_n, \theta_0)}{\partial \theta} \Big\| + o_P(1). \end{aligned} \tag{3.18}$$

Therefore, using (3.14), (3.18) implies

$$\frac{1}{\sqrt{n}}\max_{k\in\Pi_n}\Big\|\frac{\partial L_n(T_k,\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{1}{2}kF\;(\widehat{\theta}_n(T_k)-\widehat{\theta}_n(T_n))\Big\|=o_P(1). \tag{3.19}$$

Now, let $0<\tau<1$, for large value of n, we have $[\tau n]\in\Pi_n$; write

$$\begin{aligned}\frac{-2}{\sqrt{n}}G^{-1/2}\frac{\partial\widehat{L}_n(T_{[n\tau]},\widehat{\theta}_n(T_n))}{\partial\theta}&=-\frac{[n\tau]}{\sqrt{n}}G^{-1/2}F(\widehat{\theta}_n(T_{[n\tau]})-\widehat{\theta}_n(T_n))\\&\quad-2G^{-1/2}\frac{1}{\sqrt{n}}\Big[\frac{\partial\widehat{L}_n(T_{[n\tau]},\widehat{\theta}_n(T_n))}{\partial\theta}-\frac{1}{2}[n\tau]F(\widehat{\theta}_n(T_{[n\tau]})-\widehat{\theta}_n(T_n))\Big].\end{aligned}$$

Therefore, using (3.19) we have

$$-\frac{[n\tau]}{\sqrt{n}}G^{-1/2}F(\widehat{\theta}_n(T_{[n\tau]})-\widehat{\theta}_n(T_n))=\frac{-2}{\sqrt{n}}G^{-1/2}\frac{\partial\widehat{L}_n(T_{[n\tau]},\widehat{\theta}_n(T_n))}{\partial\theta}+o_P(1)$$

and the result follows by using Lemma 3.6.6. □

Proof of Theorem 3.3.2 .
Let $0<\tau_1^*<\ldots<\tau_{K^*-1}^*<1$ be the true values of breaks. Denote $t_1^*=[n\tau_1^*]$ and $t_{K-1}^*=[n\tau_{K-1}^*]$ the first and the end break instants. For n large enough , $t_1^*,t_{K-1}^*\in\Pi_n$. Therefore, it comes that, $\widehat{Q}_n^{(1)}=\max_{k\in\Pi_n}\widehat{Q}_{n,k}^{(1)}\geq\widehat{Q}_{n,t_1^*}^{(j)}$ and $\widehat{Q}_n^{(2)}=\max_{k\in\Pi_n}\widehat{Q}_{n,k}^{(2)}\geq\widehat{Q}_{n,t_{K-1}^*}^{(2)}$. Hence

$$\widehat{Q}_n=\max(\widehat{Q}_n^{(1)},\widehat{Q}_n^{(2)})\geq\max(\widehat{Q}_{n,t_1^*}^{(1)},\widehat{Q}_{n,t_{K-1}^*}^{(2)}). \tag{3.20}$$

Since $\theta_1^*,\theta_K^*\in\overset{\circ}{\Theta}(4)$, it comes from [8] that the model $\mathcal{M}_{\mathbb{Z}}(M_{\theta_1^*},f_{\theta_1^*})$ and $\mathcal{M}_{\mathbb{Z}}(M_{\theta_K^*},f_{\theta_K^*})$ have a 4-order stationary solution which we denote $(X_{t,j})_{t\in\mathbb{Z}}$ for $j=1,K$.
For $j=1,K$ denote for any $t\in\mathbb{Z}$, $q_{t,j}(\theta):=(X_{t,j}-f_\theta^{t,j})^2/(h_\theta^{t,j})+\log(h_\theta^{t,j})$ with $f_\theta^{t,j}:=f_\theta(X_{t-1,j},X_{t-2,j},\ldots)$, $h_\theta^{t,j}:=(M_\theta^{t,j})^2$ where $M_\theta^{t,j}:=M_\theta(X_{t-1,j},X_{t-2,j},\ldots)$. Also denote for $j=1,K$

$$F^{(j)}=I\!E[\frac{\partial^2 q_{0,j}(\theta_j^*)}{\partial\theta\partial\theta'}]\;\text{ and }\;G^{(j)}=I\!E\big[(\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta})(\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta})'\big].$$

Notice that, the matrix $\widehat{G}_n(T_{t_1^*})$ and $\widehat{G}_n(\overline{T}_{t_{K-1}^*})$ are symmetric positive semi-definite (by definition). For $j=1,K$, Lemma 3.6.1 implies that the matrix $G^{(j)}$ is symmetric positive definite and Corollary 5.1 of [9] implies $\widehat{G}_n(T_{t_1^*})\xrightarrow[n\to\infty]{a.s.}G^{(1)}$ and $\widehat{G}_n(\overline{T}_{t_{K-1}^*})\xrightarrow[n\to\infty]{a.s.}G^{(K)}$, hence $\widehat{G}_n(T_{t_1^*})$ and $\widehat{G}_n(\overline{T}_{t_{K-1}^*})$ are symmetric positive definite for n large enough.
According to Lemma 4 of [8], $\widehat{F}_n(T_{t_1^*})\xrightarrow[n\to\infty]{a.s.}F^{(1)}$, $\widehat{F}_n(\overline{T}_{t_{K-1}^*})\xrightarrow[n\to\infty]{a.s.}F^{(K)}$ and $F^{(1)}$, $F^{(K)}$ are non-singular. Hence, $\widehat{F}_n(T_{t_1^*})$ and $\widehat{F}_n(\overline{T}_{t_{K-1}^*})$ are non-singular for n large enough.
Therefore, the matrix $F^{(1)}(G^{(1)})^{-1}F^{(1)}$ and $F^{(K)}(G^{(K)})^{-1}F^{(K)}$ are symmetric positive definite and for n large enough, $\widehat{F}_n(T_{t_1^*})\widehat{G}_n(T_{t_1^*})^{-1}\widehat{F}_n(T_{t_1^*})$ and $\widehat{F}_n(\overline{T}_{t_{K-1}^*})\widehat{G}_n(\overline{T}_{t_{K-1}^*})^{-1}\widehat{F}_n(\overline{T}_{t_{K-1}^*})$ exist and are also symmetric positive definite. They converge almost surely to $F^{(1)}(G^{(1)})^{-1}F^{(1)}$ and $F^{(K)}(G^{(K)})^{-1}F^{(K)}$ respectively. Moreover, for n large enough and for all $U\in\mathbb{R}^d$, it holds

$$U'\widehat{\Sigma}_{n,t_1^*}U \;\geq\; \frac{t_1^*}{n}U'\widehat{F}_n(T_{t_1^*})\widehat{G}_n(T_{t_1^*})^{-1}\widehat{F}_n(T_{t_1^*})U; \tag{3.21}$$

$$U'\widehat{\Sigma}_{n,t_{K-1}^*}U \;\geq\; \frac{n-t_{K-1}^*}{n}U'\widehat{F}_n(\overline{T}_{t_{K-1}^*})\widehat{G}_n(\overline{T}_{t_{K-1}^*})^{-1}\widehat{F}_n(\overline{T}_{t_{K-1}^*})U. \tag{3.22}$$

For all $\rho > 0$ and $\theta \in \Theta$, denote $B_o(\theta,\rho)$ (rep. $B_c(\theta,\rho)$) the open (resp. closed) ball centered at θ of radius ρ in Θ. i.e.

$$B_o(\theta,\rho) = \{x \in \Theta\ ;\ \|\theta - x\| < \rho\} \ \text{ and } \ B_c(\theta,\rho) = \{x \in \Theta\ ;\ \|\theta - x\| \leq \rho\}.$$

For $A \subset \Theta$, we denote $A^c = \{x \in \Theta \ \ ; \ \ x \notin A\}$.

Since $\theta_1^* \neq \theta_K^*$ and $\theta_1^*, \theta_K^* \in \overset{\circ}{\Theta}(4) \subset \overset{\circ}{\Theta}$, then there exist $\rho_1 > 0$ and $\rho_K > 0$ such as $B_o(\theta_1^*,\rho_1) \cap B_o(\theta_K^*,\rho_K) = \emptyset$.

For all $n \in \mathbb{N}$, denote

$$\delta_n^{(1)} = \inf_{x\in B_c(\theta_1^*,\rho_1/2);\ y\in B_o^c(\theta_1^*,\rho_1)} \big((x-y)'\widehat{F}_n(T_{t_1^*})\widehat{G}_n(T_{t_1^*})^{-1}\widehat{F}_n(T_{t_1^*})(x-y)\big) \ \text{ and}$$

$$\delta_n^{(K)} = \inf_{x\in B_c(\theta_K^*,\rho_K/2);\ y\in B_o^c(\theta_K^*,\rho_K)} \big((x-y)'\widehat{F}_n(\overline{T}_{t_{K-1}^*})\widehat{G}_n(\overline{T}_{t_{K-1}^*})^{-1}\widehat{F}_n(\overline{T}_{t_{K-1}^*})(x-y)\big)$$

Also denote

$$\delta^{(j)} = \inf_{x\in B_c(\theta_j^*,\rho_j/2);\ y\in B_o^c(\theta_j^*,\rho_j)} \big((x-y)'F^{(j)}(G^{(j)})^{-1}F^{(j)}(x-y)\big) \ \text{ for } \ j=1,K.$$

It is clear that

$$\delta_n^{(j)} \xrightarrow[n\to\infty]{a.s.} \delta^{(j)} \ \text{ and } \ \delta^{(j)} > 0 \ \text{ for } \ j=1,K. \tag{3.23}$$

From [8] and [9], we have $\widehat{\theta}_n(T_{t_1^*}) \xrightarrow[n\to\infty]{a.s.} \theta_1^*$ and $\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) \xrightarrow[n\to\infty]{a.s.} \theta_K^*$. Therefore, for n large enough, $\widehat{\theta}_n(T_{t_1^*}) \in B_o(\theta_1^*,\rho_1/2)$ and $\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) \in B_o(\theta_K^*,\rho_K/2)$. Thus, two situations may occur

– if $\widehat{\theta}_n(T_n) \in B_o(\theta_K^*,\rho_K)$ i.e. $\widehat{\theta}_n(T_n) \in B_o^c(\theta_1^*,\rho_1)$ then

$$(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n))'\widehat{F}_n(T_{t_1^*})\widehat{G}_n(T_{t_1^*})^{-1}\widehat{F}_n(T_{t_1^*})(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n)) \geq \delta_n^{(1)},$$

therefore, by using (3.21) we have almost surely

$$\begin{aligned}
\widehat{Q}_{n,t_1^*}^{(1)} &:= \frac{(t_1^*)^2}{n}(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n))'\widehat{\Sigma}_{n,t_1^*}(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n)) \\
&\geq \frac{(t_1^*)^2}{n}\frac{t_1^*}{n}(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n))'\widehat{F}_n(T_{t_1^*})\widehat{G}_n(T_{t_1^*})^{-1}\widehat{F}_n(T_{t_1^*})(\widehat{\theta}_n(T_{t_1^*}) - \widehat{\theta}_n(T_n)) \\
&\geq \frac{(t_1^*)^3}{n^2}\delta_n^{(1)} \simeq n(\tau_1^*)^3\delta_n^{(1)};
\end{aligned}$$

– else $\widehat{\theta}_n(T_n) \in B_o^c(\theta_K^*,\rho_K)$ and

$$(\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) - \widehat{\theta}_n(T_n))'\widehat{F}_n(\overline{T}_{t_{K-1}^*})\widehat{G}_n(\overline{T}_{t_{K-1}^*})^{-1}\widehat{F}_n(\overline{T}_{t_{K-1}^*})(\widehat{\theta}_n(T_{t_{K-1}^*}) - \widehat{\theta}_n(T_n)) \geq \delta_n^{(K)},$$

by using (3.22) we have almost surely

$$\begin{aligned}
\widehat{Q}_{n,t_{K-1}^*}^{(2)} &:= \frac{(n-t_{K-1}^*)^2}{n}(\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) - \widehat{\theta}_n(T_n))'\widehat{\Sigma}_{n,t_{K-1}^*}(\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) - \widehat{\theta}_n(T_n)) \\
&\geq \frac{(n-t_{K-1}^*)^3}{n^2}(\widehat{\theta}_n(\overline{T}_{t_{K-1}^*}) - \widehat{\theta}_n(T_n))'\widehat{F}_n(\overline{T}_{t_{K-1}^*})\widehat{G}_n(\overline{T}_{t_{K-1}^*})^{-1}\widehat{F}_n(\overline{T}_{t_{K-1}^*}) \\
&\qquad\qquad\qquad\qquad\qquad \cdot (\widehat{\theta}_n(T_{t_{K-1}^*}) - \widehat{\theta}_n(T_n)) \\
&\geq \frac{(n-t_{K-1}^*)^3}{n^2}\delta_n^{(K)} \simeq n(1-\tau_{K-1}^*)^3\delta_n^{(K)}.
\end{aligned}$$

In all cases, it holds that $\widehat{Q}_n \geq \max(\widehat{Q}^{(1)}_{n,t^*_1}, \widehat{Q}^{(2)}_{n,t^*_{K-1}}) \geq \min\big(n(\tau^*_1)^3\delta^{(1)}_n, n(1-\tau^*_{K-1})^3\delta^{(K)}_n\big)$ *a.s.*. Thus the result follows by using (3.23). $\square$

Chapitre 4

Monitoring procedure for parameter change in causal time series

Abstract

We propose a new sequential procedure to detect change in parameter of a process $X = (X_t)_{t\in\mathbb{Z}}$ belonging to a large class of causal models. The procedure is based on a difference between the historical parameter estimate and the updated parameter estimate. Unlike classical recursive fluctuation test, the updated estimate is computed without the historical observations. These estimators are based on the quasi-likelihood of the model. The asymptotic behavior of the test is established and the consistency in power as well as an upper bound of the detection delay are obtained. Some simulation results are reported with comparisons to some other existing procedures exhibiting the accuracy of our new procedure.

Keywords : Monitoring procedure ; Change-point ; Causal processes ; Quasi-maximum likelihood estimator ; Weak convergence.

Note

The content of this chapter is based on a preprint, written in collaboration with Jean-Marc Bardet.

4.1 Introduction

In statistical inference, many authors have pointed out the danger of omitting the existence of changes in the data. Many papers have been devoted to the problem of test for parameter changes in time series models when all data are available, see for instance Horváth [32], Inclan and Tiao [37], Kokoszka and Leipus [44], Kim *et al.* [42], Aue *et al.* [8], Bardet *et al.* [9], Kengne [40]. These papers consider "retrospective" (off-line) changes *i.e.* changes in parameters when all data are available . But one might also ask what happens when new data arrive ; this is sequential change-point problem. Many papers have also focused on this problem. An important turning on this topic was made in 1996 with the works of Chu, Stinchcombe and White. They considered sequential change in regression model and pointed out the effects of repeating retrospective test every time when new data are observed ; this can increase the probability of type 1 error of the test. They successfully applied fluctuation test to solve the sequential change-point problem. Two procedures are developed based on cumulative sum (CUSUM) of residuals and recursive parameter fluctuations. Their idea has been generalized and several procedures are now based on this approach. Leisch *et al.* [56] introduced the generalized fluctuation test based on the recursive moving estimator which contains the test of Chu *et al.* [23] as a special case. Horváth *et al.* [34] introduced residual CUSUM monitoring procedure where the recursive parameter is based on the historical data. This procedure has been generalized by Aue *et al.* [2] to the class of linear model with dependent errors. Berkes *et al.* [14] considered sequential changes in the parameters of GARCH process. According to the fact that the functional limit theorem assumed by Chu *et al.* [23] is not satisfied by the squares of residuals of GARCH process, they developed a procedure based on quasi-likelihood scores. Na *et al.* [58] developed a monitoring procedure for the detection of parameter changes in general time series models. They show that under the null hypothesis of no change, their detector converges weakly to a known distribution. However, the asymptotic behavior of their detector is unknown under the alternative of parameter changes.

In this new contribution, we consider a large class of causal time series and investigate the asymptotic behavior under the alternative. More precisely, let $M, f : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ be measurable functions, $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) random variables satisfying $\mathrm{var}(\xi_0) = \sigma^2$ and let Θ be a fixed compact subset of $\mathbb{R}^d$. Let $T \subset \mathbb{Z}$, and for any $\theta \in \Theta$, define

Class $\mathcal{M}_T(M_\theta, f_\theta)$: *The process $X = (X_t)_{t\in\mathbb{Z}}$ belongs to $\mathcal{M}_T(M_\theta, f_\theta)$ if it satisfies the relation :*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{4.1}$$

The existence and properties of this general class of causal and affine processes were studied in Bardet and Wintenberger [8]. Numerous classical time series (such as AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes) are included in $\mathcal{M}_{\mathbb{Z}}(M, f)$. The off-line change detection for such a class of models has already been studied in Bardet *et al.* [9] and Kengne [40].

Suppose now that we have observed $X_1, \cdots, X_n$ which are available historical data. We assume that the historical data depends on one parameter *i.e.* there exists $\theta_0^* \in \Theta$ such as $(X_1, \cdots, X_n)$ belongs to $\mathcal{M}_{\{1,\cdots,n\}}(M_{\theta_0^*}, f_{\theta_0^*})$. Then, we observe new data $X_{n+1}, X_{n+2} \cdots, X_k, \cdots$: the monitoring starts. For each new observation, we would like to know if a change occurs in the parameter θ_0^*. More precisely, we consider the following test :

$H_0 : \theta_0^*$ is constant over the observation $X_1, \cdots, X_n, X_{n+1}, \cdots$ *i.e.* the observations $X_1, \cdots, X_n, X_{n+1}, \cdots$ belong to $\mathcal{M}_{\mathbb{N}}(M_{\theta_0^*}, f_{\theta_0^*})$;

H_1 : there exist $k^* > n$, $\theta_1^* \in \Theta$ such that $X_1, \cdots, X_n, X_{n+1}, \cdots, X_{k^*}, X_{k^*+1}, \cdots$ belongs to
$\mathcal{M}_{\{1,\cdots,k^*\}}(M_{\theta_0^*}, f_{\theta_0^*}) \bigcap \mathcal{M}_{\{k^*+1,\cdots\}}(M_{\theta_1^*}, f_{\theta_1^*})$.

When new data arrive, Chu *et al.* [23] proposed in their fluctuation procedure to compute an estimator of the parameter based on all the observations and to compare it to an estimator based on historical data. A large distance between both these estimators means that new data come from a model with different parameters. Then the null hypothesis H_0 is rejected and the monitoring stops ; otherwise, the monitoring continues. In their procedure, Leisch *et al.* [56] suggested to compute the recursive estimators on a moving window with a fixed width. They fixed a monitoring horizon so that, the procedure will stop after a fixed number of steps even if no change is detected. As Chu *et al.* [23], the recursive estimators computed by Na *et al.* [58] are based on all the observations. As we will see in the next sections, their procedure cannot be effective in terms of detection delay or to detect a small change in the parameter.

For any $k \geq 1$, $\ell, \ell' \in \{1, \cdots, k\}$ (with $\ell \leq \ell'$) let $\widehat{\theta}(X_\ell, \cdots, X_{\ell'})$ be the quasi-maximum likelihood estimator (QMLE in the sequel) of the parameter computed on $\{\ell, \cdots, \ell'\}$. When new data arrive at time $k \geq n$, we explore the segment $\{\ell, \ell+1, \cdots, k\}$ with $\ell \in \{n - v_n, n - v_n + 1 \cdots k - v_n\}$ (where $(v_n)_{n\in\mathbb{N}}$ is a fixed sequence of integer numbers) that the distance between $\widehat{\theta}(X_\ell, \cdots, X_k)$ and $\widehat{\theta}(X_1, \cdots, X_n)$ is the largest. If the norm $\|\widehat{\theta}(X_\ell, \cdots, X_k) - \widehat{\theta}(X_1, \cdots, X_n)\|$ is greater than a suitable critical value, then H_0 is rejected and the monitoring stops ; otherwise, the monitoring continues. More precisely, we construct a detector that takes into account the distance between $\widehat{\theta}(X_\ell, \cdots, X_k)$ and $\widehat{\theta}(X_1, \cdots, X_n)$. It is shown that this detector is almost surely finite under the null hypothesis and almost surely diverges to infinity under the alternative. Hence, the consistency of our procedure follows. Simulations result compared to the procedure of Horváth *et al.* [34] (see also Aue *et al.* [2]) and Na *et al.* [58] show that our procedure outperforms in terms of power and detection delay.

In the forthcoming Section 2 the assumptions and the definition of the quasi-likelihood estimator are provided. In Section 3 we present the monitoring procedure and the asymptotic results. Section 4 is devoted for a simulation study for AR(1) and GARCH$(1, 1)$ processes. The proofs of main results are presented in Section 5.

4.2 Assumptions and definition of the quasi-likelihood estimator

4.2.1 Assumptions on the class of models $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$

Let $\theta \in \mathbb{R}^d$ and M_θ and f_θ be numerical functions such that for all $(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $M_\theta\big((x_i)_{i\in\mathbb{N}}\big) \neq 0$ and $f_\theta\big((x_i)_{i\in\mathbb{N}}\big) \in \mathbb{R}$. Denote $h_\theta := M_\theta^2$. We will use the following classical notations :

1. $\|\cdot\|$ applied to a vector denotes the Euclidean norm of the vector ;
2. for any compact set $\mathcal{K} \subseteq \mathbb{R}^d$ and for any $g : \mathcal{K} \longrightarrow \mathbb{R}^{d'}$, $\|g\|_{\mathcal{K}} = \sup_{\theta\in\mathcal{K}}(\|g(\theta)\|)$;

3. for any set $\mathcal{K} \subseteq \mathbb{R}^d$, $\overset{\circ}{\mathcal{K}}$ denotes the interior of $\mathcal{K}$.

Throughout the sequel, we will assume that the functions $\theta \mapsto M_\theta$ and $\theta \mapsto f_\theta$ are twice continuously differentiable on Θ. Let $\Psi_\theta = f_\theta$, M_θ and $i = 0, 1, 2$, then for any compact set $\mathcal{K} \subset \Theta$ define

Assumption $\mathbf{A}_i(\Psi_\theta, \mathcal{K})$: *Assume that $\|\partial^i \Psi_\theta(0)/\partial\theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real numbers $(\alpha_j^{(i)}(\Psi_\theta, \mathcal{K}))_{j\geq 1}$ such that $\sum\limits_{j=1}^{\infty} \alpha_j^{(i)}(\Psi_\theta, |) < \infty$ and*

$$\Big\|\frac{\partial^i \Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i \Psi_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{j=1}^{\infty} \alpha_j^{(i)}(\Psi_\theta, \mathcal{K})|x_j - y_j| \quad \textit{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

In the sequel we refer to the particular case called "ARCH-type process" if $f_\theta = 0$ and if the following assumption holds with $h_\theta = M_\theta^2$:

Assumption $\mathbf{A}_i(h_\theta, \mathcal{K})$: *Assume that $\|\partial^i h_\theta(0)/\partial\theta^i\|_\Theta < \infty$ and there exists a sequence of non-negative real numbers $(\alpha_j^{(i)}(h_\theta, \mathcal{K}))_{j\geq 1}$ such as $\sum\limits_{j=1}^{\infty} \alpha_j^{(i)}(h_\theta, \mathcal{K}) < \infty$ and*

$$\Big\|\frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i}\Big\|_{\mathcal{K}} \leq \sum_{j=1}^{\infty} \alpha_j^{(i)}(h_\theta, \mathcal{K})|x_j^2 - y_j^2| \quad \textit{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

The Lipschitz-type hypothesis $A_i(\Psi_\theta, \mathcal{K})$ are classical when studying the existence of solutions of the general model (see for instance [27]). Using a result of [8], for each model $\mathcal{M}_{\mathbb{Z}}(M_\theta, f_\theta)$ it is interesting to define the following set :

$$\begin{aligned}\Theta(r) := \Big\{\theta \in \Theta,\ A_0(f_\theta, \{\theta\}) \text{ and } A_0(M_\theta, \{\theta\}) \text{ hold with } &\sum_{j\geq 1} \alpha_j^{(0)}(f_\theta, \{\theta\}) \\ &+ (\mathbb{E}|\xi_0|^r)^{1/r} \sum_{j\geq 1} \alpha_j^{(0)}(M_\theta, \{\theta\}) < 1\Big\} \\ \bigcup \Big\{\theta \in \Theta,\ f_\theta = 0 \text{ and } A_0(h_\theta, \{\theta\}) \text{ holds with } (\mathbb{E}|\xi_0|^r)^{2/r} &\sum_{j\geq 1} \alpha_j^{(0)}(h_\theta, \{\theta\}) < 1\Big\}.\end{aligned}$$

Then, if $\theta \in \Theta(r)$ the existence of a unique causal, stationary and ergodic solution $X = (X_t)_{t\in\mathbb{Z}} \in \mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$ is ensured (see more details in [8]). The subset $\Theta(r)$ is defined as a reunion to consider accurately general causal models and ARCH-type models simultaneously.

Here there are assumptions required for studying QLME asymptotic properties :

Assumption $\mathbf{D}(\Theta)$: $\exists \underline{h} > 0$ such that $\inf\limits_{\theta\in\Theta}(|h_\theta(x)|) \geq \underline{h}$ for all $x \in \mathbb{R}^{\mathbb{N}}$.

Assumption Id(Θ) : For all $(\theta, \theta') \in \Theta^2$,

$$\Big(f_\theta(X_0, X_{-1}, \cdots) = f_{\theta'}(X_0, X_{-1}, \cdots) \text{ and } h_\theta(X_0, X_{-1}, \cdots) = h_{\theta'}(X_0, X_{-1}, \cdots) \text{ a.s.}\Big) \Rightarrow\ \theta = \theta'.$$

Assumption Var(Θ) : For all $\theta \in \Theta$, one of the families $\big(\dfrac{\partial f_\theta}{\partial\theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ or $\big(\dfrac{\partial h_\theta}{\partial\theta^i}(X_0, X_{-1}, \cdots)\big)_{1\leq i\leq d}$ is a.s. linearly independent.

Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$: for i= 0, 1, 2, $\mathbf{A}_i(f_\theta, \Theta)$ and $\mathbf{A}_i(M_\theta, \Theta)$ (or $\mathbf{A}_i(h_\theta, \Theta)$) hold and there exists $\ell > 2$ such that $\alpha_j^{(i)}(f_\theta, \Theta) + \alpha_j^{(i)}(M_\theta, \Theta) + \alpha_j^{(i)}(h_\theta, \Theta) = \mathcal{O}(j^{-\ell})$ for $j \in \mathbb{N}$.

Note that in this last assumption, as in [8], we use the convention that if $\mathbf{A}_i(M_\theta, \Theta)$ holds then $\alpha_\ell^{(i)}(h_\theta, \Theta) = 0$ and if $\mathbf{A}_i(h_\theta, \Theta)$ holds then $\alpha_\ell^{(i)}(M_\theta, \Theta) = 0$.

4.2.2 Two first examples

1. ARMA(p,q) processes.
   Consider the ARMA(p,q) process defined by :
   $$X_t + \sum_{i=1}^{p} a_i^* X_{t-i} = \sum_{j=0}^{q} b_j^* \xi_{t-j} \ , \ t \in \mathbb{Z} \tag{4.2}$$
   with $b_0^* \neq 0$, $\theta_0^* = (a_1^*, \cdots, a_p^*, b_0^*, \cdots, b_q^*) \in \Theta \subset \mathbb{R}^{p+q+1}$ and (ξ_t) a white noise such as $I\!E(\xi_0^2) = 1$. When $\sum_{j=0}^q b_j^* X^j \neq 0$ and $1 + \sum_{i=0}^p a_i^* X^i \neq 0$ for all $|X| \leq 1$, this process can be also written as :
   $$X_t = b_0^* \xi_t + \sum_{j=1}^{\infty} \phi_j(\theta_0^*)\, X_{t-i} \ , \ t \in \mathbb{Z}$$
   where $\theta \in \Theta \mapsto \phi_j(\theta)$ are functions only depending on θ and decreasing exponentially fast to 0 ($j \to \infty$). The process (4.2) belongs to the class $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ where $f_\theta(x_1, \cdots) = \sum_{j\geq 1} \phi_j(\theta) x_j$ and $M_\theta \equiv b_0^*$ for all $\theta \in \Theta$. Then Assumptions $\mathbf{D}(\Theta)$, $\mathbf{A}_0(f_\theta, \Theta)$, $\mathbf{A}_0(M_\theta, \Theta)$ hold with $\underline{h} = |b_0^*| > 0$ and $\alpha_j^{(0)}(f_\theta, \Theta) = \|\phi_j(\theta)\|_\Theta$ while $\alpha_j^{(0)}(M_\theta, \Theta) = 0$ for $j \in \mathbb{N}^*$. Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds since there exists $c > 0$ and $C > 0$ such as $|\phi_j| \leq C\, e^{-cj}$ for $j \in \mathbb{N}$. Moreover, if (ξ_t) is a sequence of non-degenerate random variables (*i.e.* ξ_t are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold. Finally, for any $r \geq 1$ such that $I\!E|\xi_0|^r < \infty$, then $\Theta(r) = \{\theta \in \mathbb{R}^{p+q+1}, \ \sum_{j\geq 1} |\phi_j(\theta)| < 1\}$. Note that if $\theta \in \Theta(r)$ with $r \geq 1$ then the previous conditions of stationarity $\sum_{j=0}^q b_j X^j \neq 0$ and $1 + \sum_{i=0}^p a_i X^i \neq 0$ for all $|X| \leq 1$ are satisfied.

2. GARCH(p,q) processes.
   Consider the GARCH(p,q) process defined by :
   $$X_t = \sigma_t\, \xi_t \ , \ \sigma_t^2 = a_0^* + \sum_{j=1}^{p} a_j^* X_{t-j}^2 + \sum_{j=1}^{q} b_j^* \sigma_{t-j}^2 \ , \quad t \in \mathbb{Z} \tag{4.3}$$
   with $I\!E(\xi_0^2) = 1$ and $\theta_0^* := (a_0^*, \cdots, a_p^*, b_1^*, \cdots, b_q^*) \in \Theta$ where Θ is a compact subset of $]0, \infty[\times [0, \infty[^{p+q}$ such that $\sum_{j=1}^p a_j + \sum_{j=1}^q b_j < 1$ for all $\theta \in \Theta$. Then there exists (see Bollerslev [20] or Nelson and Cao [60]) a nonnegative sequence $(\psi_j(\theta_0^*))_{j\geq 0}$ such that $\sigma_t^2 = \psi_0(\theta_0^*) + \sum_{j\geq 1} \psi_j(\theta_0^*) X_{t-j}^2$ with $\psi_0(\theta_0^*) = a_0^*/(1 - \sum_{j=1}^q b_j^*)$.
   This process belongs to the class $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*}, f_{\theta_0^*})$ where $f_\theta \equiv 0$ and $M_\theta(x_1, \cdots) = \sqrt{\psi_0(\theta) + \sum_{j\geq 1} \psi_j(\theta) x_j^2}$ for all $\theta \in \Theta$. Assumption $\mathbf{D}(\Theta)$ holds with $\underline{h} = \inf_{\theta\in\Theta}(\psi_0(\theta)) > 0$. If there exists $0 < \rho_0 < 1$ such that for any $\theta \in \Theta$, $\sum_{j=1}^q a_j + \sum_{j=1}^p b_j \leq \rho_0$ then the sequences $(\|\psi_j(\theta)\|_\Theta)_{j\geq 1}$, $(\|\psi_j'(\theta)\|_\Theta)_{j\geq 1}$ and $(\|\psi_j''(\theta)\|_\Theta)_{j\geq 1}$ decay exponentially fast (see Berkes *et al.* [13]) and Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds. Moreover, (ξ_t^2) is a

sequence of non-degenerate random variables (*i.e.* ξ_t^2 are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold. Finally for $r \geq 2$ we obtain

$$\Theta(r) = \{\theta \in \Theta \ ; \ (\mathbb{E}|\xi_0|^r)^{2/r} \sum_{j=1}^{\infty} \phi_j(\theta) < 1\}.$$

4.2.3 The quasi-maximum likelihood estimator

Let $k \geq n \geq 2$, if $(X_1, \cdots, X_k) \in \mathcal{M}_{\{1,\cdots,k\}}(M_\theta, f_\theta)$, then for $T \subset \{1, \cdots, k\}$, the conditional quasi-(log)likelihood computed on T is given by :

$$L(T, \theta) := -\frac{1}{2} \sum_{t \in T} q_t(\theta) \quad \text{with} \quad q_t(\theta) = \frac{(X_t - f_\theta^t)^2}{h_\theta^t} + \log(h_\theta^t) \tag{4.4}$$

where $f_\theta^t = f_\theta(X_{t-1}, X_{t-2} \ldots)$, $M_\theta^t = M_\theta(X_{t-1}, X_{t-2} \ldots)$ and $h_\theta^t = M_\theta^{t2}$. The classical approximation of this conditional log-likelihood is given by :

$$\widehat{L}(T, \theta) := -\frac{1}{2} \sum_{t \in T} \widehat{q}_t(\theta) \quad \text{where} \quad \widehat{q}_t(\theta) := \frac{\big(X_t - \widehat{f}_\theta^t\big)^2}{\widehat{h}_\theta^t} + \log\big(\widehat{h}_\theta^t\big) \tag{4.5}$$

with $\widehat{f}_\theta^t = f_\theta(X_{t-1}, \ldots, X_1, 0, 0, \cdots)$, $\widehat{M}_\theta^t = M_\theta(X_{t-1}, \ldots, X_1, 0, 0, \cdots)$ and $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$. For $T \subset \{1, \cdots, k\}$, define the quasi maximum-likelihood estimator (QMLE) of θ computed on T by

$$\widehat{\theta}(T) := \underset{\theta \in \Theta}{\operatorname{argmax}}(\widehat{L}(T, \theta)). \tag{4.6}$$

In Bardet and Wintenberger [8] it was established that if $(X_1, \cdots, X_n)$ is an observed trajectory of $X \subset \mathcal{M}_{\mathbb{Z}}(f_{\theta_0^*}, M_{\theta_0^*})$ with $\theta_0^* \in \overset{\circ}{\Theta}(4)$ and if Θ is a compact set such as Assumptions $\mathbf{A}_i(f_\theta, M_\theta, \Theta)$ (or $\mathbf{A}_i(h_\theta, \Theta)$) hold for $i = 0, 1, 2$ and under Assumptions $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$, $\mathbf{Var}(\Theta)$, $\mathbf{K}(f_\theta, M_\theta, \Theta)$, then

$$\sqrt{n}\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0\,,\, F\, G^{-1}\, F), \tag{4.7}$$

with

$$G := \mathbb{E}\Big[\frac{\partial q_0(\theta_0^*)}{\partial \theta} \frac{\partial q_0(\theta_0^*)}{\partial \theta}'\Big] \quad \text{and} \quad F := \mathbb{E}\Big[\frac{\partial^2 q_0(\theta_0^*)}{\partial \theta \partial \theta'}\Big], \tag{4.8}$$

where $'$ denotes the transpose and with q_0 defined in (4.4). Note that under assumptions $\mathbf{D}(\Theta)$ and $\mathbf{Var}(\Theta)$, G is symmetric positive definite (see [40]) and F is non-singular (see [8]). Also define the matrix

$$\widehat{G}(T) := \frac{1}{\mathrm{Card}(T)} \sum_{t \in T} \Big(\frac{\partial \widehat{q}_t(\widehat{\theta}(T))}{\partial \theta}\Big)\Big(\frac{\partial \widehat{q}_t(\widehat{\theta}(T))}{\partial \theta}\Big)' \quad \text{and} \quad \widehat{F}(T) := -\frac{2}{\mathrm{Card}(T)} \Big(\frac{\partial^2 \widehat{L}_m(T, \widehat{\theta}(T))}{\partial \theta \partial \theta'}\Big). \tag{4.9}$$

Under the previous assumptions, $\widehat{G}(T_{1,n})$ and $\widehat{F}(T_{1,n})$ respectively converge almost surely to G and F respectively. Hence, $\sqrt{n}\, \widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) \xrightarrow[n \to \infty]{\mathcal{D}} \mathcal{N}(0\,,\, I_d)$ with I_d the identity matrix. This result will be the starting point of the following monitoring procedure.

4.3 The monitoring procedure and asymptotic results

4.3.1 The monitoring procedure

In the sequel, $(X_1, \cdots, X_n)$ is supposed to be the historical available observations belongs to the class $\mathcal{M}_{\{1,\cdots,n\}}(f_{\theta_0^*}, M_{\theta_0^*})$. For $1 \leq \ell \leq \ell'$, denote

$$T_{\ell,\ell'} := \{\ell, \ell+1, \cdots, \ell'\}.$$

At a monitoring instant k, our idea is to evaluate the difference between $\widehat{\theta}(T_{\ell,k})$ and $\widehat{\theta}(T_{1,n})$ for any $\ell = n, \cdots, k$. More precisely, for any $k > n$ define the statistic (called the detector)

$$\widehat{C}_{k,\ell} := \sqrt{n}\,\frac{k-\ell}{k}\,\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})\big)\big\|$$

for $\ell = n, \cdots, k$. Since the matrix $\widehat{G}(T_{1,n})$ is asymptotically symmetric and positive definite (see [40]), $\widehat{G}(T_{1,n})^{-1/2}$ exists for n large enough and $\widehat{C}_{k,\ell}$ is well defined. At the beginning of the monitoring and when ℓ is close to k, the length of $T_{\ell,k}$ is too small, therefore the numerical algorithm used to compute $\widehat{\theta}(T_{\ell,k})$ cannot converge. This can introduce a large distortion in the procedure. To avoid this, we introduce an integer sequence $(v_n)_{n\in\mathbb{N}}$ with $v_n \leq n$ and compute $\widehat{C}_{k,\ell}$ for $\ell \in \{n - v_n, n - v_n + 1, \cdots, k - v_n\}$. Thus, for any $k > n$ denote

$$\Pi_{n,k} := \{n - v_n, n - v_n + 1, \cdots, k - v_n\}.$$

For technical reasons, assume that,

$$v_n \to \infty \quad \text{and} \quad v_n/\sqrt{n} \to 0 \;\; (n \to \infty).$$

According to Remark 2.1 of [40], we can choose $v_n = [(\log n)^\delta]$ with $\delta > 1$.
Note that, if change does not occur at time $k > n$, for any $\ell \in \Pi_{n,k}$, the two estimators $\widehat{\theta}(T_{\ell,k})$ and $\widehat{\theta}(T_{1,n})$ are close and the detector $\widehat{C}_{k,\ell}$ is not large enough. Hence, the monitoring stops and reject H_0 at the first time $k > n$ where there exists $\ell \in \Pi_{n,k}$ satisfying $\widehat{C}_{k,\ell} > c$ for a fixed constant $c > 0$. To be more general, we will use a $b : (0,\infty) \mapsto (0,\infty)$, called a boundary function satisfying :

Assumption B : $b : (0,\infty) \mapsto (0,\infty)$ is a non-increasing and continuous function such as $\underset{0<t<\infty}{\mathrm{Inf}}\, b(t) > 0$.

Then the monitoring procedure stops at the first time $k > n$ such as there exists $\ell \in \Pi_{n,k}$ satisfying $\widehat{C}_{k,\ell} > b((k-\ell)/n)$. Hence define the stopping time :

$$\tau(n) := \mathrm{Inf}\{k > n \,/\, \exists \ell \in \Pi_{n,k},\ \widehat{C}_{k,\ell} > b((k-\ell)/n)\} = \mathrm{Inf}\Big\{k > n \,/\, \max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\}.$$

Therefore, we have

$$P\{\tau(n) < \infty\} = P\Big\{\max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1 \text{ for some } k > n\Big\} = P\Big\{\sup_{k>n}\max_{\ell\in\Pi_{n,k}} \frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)} > 1\Big\}. \tag{4.10}$$

The challenge is to choose a suitable boundary function $b(\cdot)$ such as for some given $\alpha \in (0,1)$

$$\lim_{n\to\infty} P_{H_0}\{\tau(n) < \infty\} = \alpha$$

and
$$\lim_{n\to\infty} P_{H_1}\{\tau(n) < \infty\} = 1$$
where the hypothesis H_0 and H_1 are specified in Section 4.1. In the case where $b(\cdot)$ is a constant positive value, $b \equiv c$ with $c > 0$, these conditions lead to compute a threshold $c = c_\alpha$ depending on α. If change is detected under H_1 i.e. $\tau(n) < \infty$ and $\tau(n) > k^*$, then the detection delay is defined by $\widehat{d}_n = \tau(n) - k^*$.

Using the previous notations, Na *et al.* [58] used the following detector
$$\widehat{D}_k := \sqrt{n}\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{1,n})\big)\|.$$
At the step k of the monitoring, their recursive estimator is based on $X_1, \cdots, X_n, \cdots, X_k$. One can see that this estimator is highly influenced by the historical data. Assume that a change occurs at time $k^* \le k$, in the sequel of the procedure, the recursive estimator contents is based on $X_1, \cdots, X_n, \cdots, X_{k^*-1}$ which depends on θ_0^*. Then, one must wait longer before the difference between $\widehat{\theta}(X_1, \cdots, X_n)$ and $\widehat{\theta}(X_1, \cdots, X_n, \cdots, X_k)$ been significant at a step $k > k^*$. Therefore, their procedure cannot be effective in terms of detection delay. Moreover, if n is carried to infinity, it is not sure that this change will be detected. These are confirmed by the results of simulations (see Section 4).

For the sequential change in GARCH(p,q) models, Berkes *et al.* (2004) considered an estimator based on historical data to compute quasi-likelihood scores. They used the fact that the partial derivatives applied to a vector $\mathbf{u}$ is equal to 0 if and only if $\mathbf{u}$ is the true parameter of the model. So, when change occurs, their detector growths asymptotically to infinity. Therefore, their procedure is consistent. But, it is not sure that this result can be extended to a large class of model as we have considered here.

4.3.2 Asymptotic behaviour under the null hypothesis

Under H_0, the parameter θ_0^* does not change over the new observations. Thus we have the result

Theorem 4.3.1. Assume $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$, $\mathbf{Var}(\Theta)$, $\mathbf{K}(f_\theta, M_\theta, \Theta)$, $\mathbf{B}$ and $\theta_0^* \in \overset{\circ}{\Theta}(4)$. Under null hypothesis H_0, then
$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1}\sup_{1<s<t} \frac{\|W_d(s) - sW_d(1))\|}{t\ b(s)} > 1\Big\}$$
where W_d is a d-dimensional standard Brownian motion.

In the simulations, we will use the most "natural" boundary function $b(\cdot) = c$ with c a positive constant since it satisfies the above assumptions imposed to $b(\cdot)$. In such a case, the forthcoming corollary indicates that the asymptotic distribution of Theorem 4.3.1 can be easily computed :

Corollary 4.3.1. Assume $b(t) = c > 0$ for $t \ge 0$. Under the assumptions of Theorem 4.3.1,
$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{\sup_{t>1}\sup_{1<s<t} \frac{1}{t}\|W_d(s) - sW_d(1))\| > c\Big\} = P\{U_d > c\}$$
where $U_d = \sup_{0<u<1} f(u)\,\|W_d(u)\|$ with $f(u) = \dfrac{\sqrt{9-u}+\sqrt{1-u}}{\sqrt{9-u}+3\sqrt{1-u}}\Big(\dfrac{2}{3-u+\sqrt{(9-u)(1-u)}}\Big)^{1/2}$.

Therefore, at a nominal level $\alpha \in (0,1)$, take $c = c(\alpha)$ be the $(1-\alpha)$-quantile of the distribution of U_d. Table 4.1 shows the $(1-\alpha)$-quantile of this distribution for $\alpha = 0.01, 0.05, 0.10$ and $d = 1, \cdots, 5$ computed through Monte-Carlo simulations.

	$d=1$	$d=2$	$d=3$	$d=4$	$d=5$
$\alpha=0.01$	2.583	3.035	3.335	3.631	3.914
$\alpha=0.05$	1.954	2.432	2.760	3.073	3.334
$\alpha=0.10$	1.652	2.156	2.486	2.784	3.028

TABLE 4.1 – Empirical $(1-\alpha)$-quantile of the distribution of U_d, for $d=1,\cdots,5$.

4.3.3 Asymptotic under the alternative

Under the alternative H_1, the parameter changes from θ_0^* to θ_1^* at $k^*>n$, where $\theta_1^*\in\Theta$ and $\theta_0^*\neq\theta_1^*$. Then

Theorem 4.3.2. Assume $\mathbf{D}(\Theta)$, $\mathbf{Id}(\Theta)$,$\mathbf{Var}(\Theta)$, $\mathbf{K}(f_\theta,M_\theta,\Theta)$ and $\mathbf{B}$. Under the alternative H_1, if $\theta_1^*\neq\theta_0^*$ and $\theta_0^*,\theta_1^*\in\overset{\circ}{\Theta}(4)$ then for $k^*=k^*(n)$ such as $\limsup_{n\to\infty}k^*(n)/n<\infty$ and $k_n=k^*(n)+n^\delta$ with $\delta\in(1/2,1)$,

$$\max_{\ell\in\Pi_{n,k_n}}\frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)}\xrightarrow[n\to\infty]{a.s.}\infty.$$

The forthcoming Corollary 4.3.2 can be immediately deduced from the relation (4.10).

Corollary 4.3.2. Under assumptions of Theorem 4.3.2,

$$\lim_{n\to\infty}P\{\tau(n)<\infty\}=1.$$

Remark 4.3.1. We know that the monitoring is stopped and rejects H_0 at the first time k when

$$\max_{\ell\in\Pi_{n,k}}\frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)}>1.$$

Therefore, it follows from Theorem 4.3.2 that under the hypothesis H_1, the detection delay $\widehat{d}_n$ of the procedure can be bounded by $\mathcal{O}(n^{1/2+\varepsilon})$ for any $\varepsilon>0$ (or even by $\mathcal{O}\big(\sqrt{n}(\log n)^a\big)$ with $a>0$ using the same kind of proof).

4.3.4 Examples

AR(∞) processes

Consider the generalization of ARMA(p,q) processes defined in (4.2) *i.e.* a AR(∞) processes defined by :

$$X_t=\phi_0(\theta_0^*)+\sum_{j\geq1}\phi_j(\theta_0^*)X_{t-j}+\xi_t\ ,\ t\in\mathbb{Z}\tag{4.11}$$

with $\theta_0^*\in\overset{\circ}{\Theta}$, where we can chose Θ as a compact subset of $\Theta(4)\subset\mathbb{R}^d$ where

$$\Theta(4)=\{\theta\in\mathbb{R}^d;\ \sum_{j\geq1}|\phi_j(\theta)|<1\}.$$

This process belongs to the class $\mathcal{M}_\mathbb{Z}(M_{\theta_0^*},f_{\theta_0^*})$ where $f_\theta(x_1,\cdots)=\sum_{j\geq1}\phi_j(\theta)x_j$ and $M_\theta\equiv\phi_0(\theta)$ for all $\theta\in\Theta$ and therefore $\alpha_j^{(0)}(f_\theta,\Theta)=\|\phi_j(\theta)\|_\Theta$ and $\alpha_j^{(0)}(M_\theta,\Theta)=0$ for $j\in\mathbb{N}^*$. Then

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta\in\Theta}(|\phi_0(\theta)|) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto \phi_j(\theta)$ are twice differentiable functions satisfying $\max\big(\|\psi_j(\theta)\|_\Theta, \|\phi'_j(\theta)\|_\Theta, \|\phi''_j(\theta)\|_\Theta\big) = O(j^{-\ell})$ for $j \in \mathbb{N}$.
- if (ξ_t) is a sequence of non-degenerate random variables (*i.e.* ξ_t are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold.

Case of AR(p) process
Assume that

$$X_t = \phi_0^* + \sum_{j=1}^{p} \phi_j^* X_{t-j} + \xi_t \quad \text{with } p \in \mathbb{N}^*.$$

The true parameter of the model is denoted by $\theta_0^* = (\phi_0^*, \phi_1^*, \cdots, \phi_p^*) \in \Theta$ where $\Theta = \{\theta = (\phi_0, \phi_1, \cdots, \phi_p) \in \mathbb{R}^{p+1} \ \ / \sum_{j=1}^{p} |\phi_j| < 1\}$. Then, $\Theta(r) = \Theta$ for any $r \geq 1$. Assume that a trajectory $(X_1, \cdots, X_k)$ has been observed, for any $t = 1, \cdots, k$ and $\theta \in \Theta$ we have, $\widehat{q}_t(\theta) = \big(X_t - \phi_0 - \sum_{j=1}^{p} \phi_j X_{t-j}\big)^2$, $\dfrac{\partial \widehat{q}_t(\theta)}{\partial\theta} = -2\big(X_t - \phi_0 - \sum_{j=1}^{p} \phi_j X_{t-j}\big) \cdot (1, X_{t-1}, X_{t-2}, \cdots, X_{t-p})$. Moreover, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial\phi_0\partial\phi_0} = 2$, for $j = 1, \cdots, p$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial\phi_0\partial\phi_j} = 2X_{t-j}$ and for $1 \leq i, j \leq p$, $\dfrac{\partial^2 \widehat{q}_t(\theta)}{\partial\phi_i\partial\phi_j} = 2X_{t-i}X_{t-j}$.

ARCH(∞) processes

Consider the generalization GARCH(p, q) processes defined in (4.3) *i.e.* a ARCH(∞) processes defined by :

$$X_t = \sigma_t\, \xi_t \quad \text{and} \quad \sigma_t^2 = \psi_0(\theta_0^*) + \sum_{j=1}^{\infty} \psi_j(\theta_0^*) X_{t-j}^2\ ,\ t \in \mathbb{Z} \tag{4.12}$$

with $\theta_0^* \in \overset{\circ}{\Theta}$, where we can chose Θ as a compact subset of $\Theta(4) \subset \mathbb{R}^d$ where

$$\Theta(4) = \{\theta \in \mathbb{R}^d;\ (\mathbb{E}|\xi_0|^4)^{1/2} \sum_{j=1}^{\infty} |\phi_j(\theta)| < 1\}.$$

This process, introduced by Robinson [64], belongs to the class $\mathcal{M}_{\mathbb{Z}}(f_{\theta_0^*}, M_{\theta_0^*},)$ where $f_\theta(x_1, \cdots) \equiv 0$ and $M_\theta^2(x_1, \cdots) = \psi_0(\theta) + \sum_{j\geq 1} \psi_j(\theta) x_j^2$ for all $\theta \in \Theta$ and therefore $\alpha_j^{(0)}(f_\theta, \Theta) = 0$ and $\alpha_j^{(0)}(h_\theta, \Theta) = \|\phi_j(\theta)\|_\Theta$ for $j \in \mathbb{N}^*$ (X is of course a ARCH-type process). Then

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta\in\Theta}(\psi_0(\theta)) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto \phi_j(\theta)$ are twice differentiable functions satisfying $\max\big(\|\psi_j(\theta)\|_\Theta, \|\psi'_j(\theta)\|_\Theta, \|\psi''_j(\theta)\|_\Theta\big) = O(j^{-\ell})$ for $j \in \mathbb{N}$.
- if (ξ_t^2) is a sequence of non-degenerate random variables (*i.e.* ξ_t^2 are not equal to a constant), Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$ hold.

Case of GARCH$(1, 1)$ process
Assume that

$$X_t = \sigma_t \xi_t \quad \text{with } \ \sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$$

with $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*) \in \Theta \subset]0,\infty[\times[0,\infty[^2$ and satisfying $\alpha_1^* + \beta_1^* < 1$. The ARCH($\infty$) representation is $\sigma_t^2 = \alpha_0^*/(1-\beta_1^*) + \alpha_1^* \sum_{j\geq 1} (\beta_1^*)^{j-1} X_{t-j}^2$. If a trajectory $(X_1, \cdots, X_k)$ has been observed, for any $t = 1, \cdots, k$ and $\theta \in \Theta$ we have,

$$\widehat{h}_\theta^t = \alpha_0/(1-\beta_1) + \alpha_1 X_{t-1}^2 + \alpha_1 \sum_{j=2}^t \beta_1^{j-1} X_{t-j}^2 \quad \text{and} \quad \widehat{q}_t(\theta) = X_t^2 / \ \widehat{h}_\theta^t + \log(\widehat{h}_\theta^t).$$

Therefore, it follows that $\dfrac{\partial \widehat{q}_t(\theta)}{\partial\theta} = \dfrac{1}{\widehat{h}_\theta^t}\Big(1 - \dfrac{X_t^2}{\widehat{h}_\theta^t}\Big)\Big(\dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_0}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \alpha_1}, \dfrac{\partial \widehat{h}_\theta^t}{\partial \beta_1}\Big)$ with $\partial\widehat{h}_\theta^t/\partial\alpha_1 = X_{t-1}^2 + \sum_{j=2}^t \beta_1^{j-1} X_{t-j}^2$ $\partial\widehat{h}_\theta^t/\partial\alpha_0 = 1/(1-\beta_1)$, and $\partial\widehat{h}_\theta^t/\partial\beta_1 = \alpha_0/(1-\beta_1)^2 + \alpha_1 X_{t-2}^2 + \alpha_1 \sum_{j=3}^t (j-1)\beta_1^{j-2} X_{t-k}^2$.
Let $\theta = (\alpha_0, \alpha_1, \beta_1) = (\theta_1, \theta_2, \theta_3) \in \Theta$, for $1 \leq i, j \leq 3$, we have

$$\frac{\partial^2 \widehat{q}_t(\theta)}{\partial\theta_i\partial\theta_j} = \frac{1}{(\widehat{h}_\theta^t)^2}\Big(\frac{2X_t^2}{\widehat{h}_\theta^t} - 1\Big)\frac{\partial \widehat{h}_\theta^t}{\partial\theta_i}\frac{\partial \widehat{h}_\theta^t}{\partial\theta_j} + \frac{1}{\widehat{h}_\theta^t}\Big(1 - \frac{X_t^2}{\widehat{h}_\theta^t}\Big)\frac{\partial^2 \widehat{h}_\theta^t}{\partial\theta_i\partial\theta_j}$$

with $\partial^2\widehat{h}_\theta^t/\partial\alpha_0^2 = 0$, $\partial^2\widehat{h}_\theta^t/\partial\alpha_0\partial\alpha_1 = 0$, $\partial^2\widehat{h}_\theta^t/\partial\alpha_1^2 = 0$, $\partial^2\widehat{h}_\theta^t/\partial\alpha_1\partial\beta_1 = X_{t-2}^2 + \sum_{j=3}^t (j-1)\beta_1^{j-2} X_{t-j}^2$, $\partial^2\widehat{h}_\theta^t/\partial\alpha_0\partial\beta_1 = 1/(1-\beta_1)^2$ and $\partial\widehat{h}_\theta^t/\partial\beta_1^2 = 2\alpha_0/(1-\beta_1)^3 + 2\alpha_1 X_{t-3}^2 + \alpha_1 \sum_{j=4}^t (j-1)(j-2)\beta_1^{j-3} X_{t-j}^2$.

TARCH(∞) processes

The process X is called Threshold ARCH(∞) (TARCH(∞) in the sequel) if it satisfies

$$X_t = \sigma_t \xi_t \quad \text{and} \quad \sigma_t = b_0(\theta_0^*) + \sum_{j=1}^\infty \Big[b_j^+(\theta_0^*) \max(X_{t-j}, 0) - b_j^-(\theta_0^*) \min(X_{t-j}, 0)\Big]\ , \ t \in \mathbb{Z} \tag{4.13}$$

where the parameters $b_0(\theta)$, $b_j^+(\theta)$ and $b_j^-(\theta)$ are assumed to be non negative real numbers and $\theta \in \overset{\circ}{\Theta}$ where Θ is a compact subset of $\Theta(4)$ where

$$\Theta(4) = \Big\{\theta \in \mathbb{R}^d \Big/ \left(\mathbb{E}|\xi_0|^4\right)^{1/4} \sum_{j=1}^\infty \max\left(b_j^-(\theta), b_j^+(\theta)\right) < 1\Big\}$$

since $\alpha_j^{(0)}(M, \{\theta\}) = \max\left(b_j^-(\theta), b_j^+(\theta)\right)$. This class of processes is a generalization of the class of TGARCH(p,q) processes (introduced by Rabemananjara and Zakoïan [62]). Then,

- Assumption $\mathbf{D}(\Theta)$ holds if $\underline{h} = \inf_{\theta\in\Theta} b_0(\theta) > 0$;
- Assumption $\mathbf{K}(f_\theta, M_\theta, \Theta)$ holds if there exists $\ell > 2$ and and if $\theta \mapsto b_j^-(\theta)$ and $\theta \mapsto b_j^+(\theta)$ are twice differentiable functions and satisfying for all $j \in \mathbb{N}$

$$\max\left(\|b_j^-(\theta)\|_\Theta, \|b_j^+(\theta)\|_\Theta, \|\frac{\partial}{\partial\theta} b_j^-(\theta)\|_\Theta, \|\frac{\partial}{\partial\theta} b_j^+(\theta)\|_\Theta, \|\frac{\partial^2}{\partial\theta^2} b_j^-(\theta)\|_\Theta, \|\frac{\partial^2}{\partial\theta^2} b_j^+(\theta)\|_\Theta\right) = O(j^{-\ell}).$$

Unfortunately, for TARCH(∞) it is not possible to provide as for AR(∞) or ARCH(∞) processes simple conditions for obtaining Assumptions $\mathbf{Id}(\Theta)$ and $\mathbf{Var}(\Theta)$.

4.4 Some simulation and numerical experiments

First remark that, at a time $k > n$, we need to compute $\widehat{C}_{k,\ell}$ for all $\ell \in \Pi_{n,k}$ to test whether change occurs or not. On can see that, the computational time is very long and increase with k. To reduce this, we introduce an integer sequence (u_n) (satisfying $u_n/\sqrt{n} \to 0$ as $n \to \infty$; typically $u_n = [\ln(n)]$) and compute $\widehat{C}_{k,\ell}$ only for

$$\ell \in \Pi^0_{n,k} := \{n - v_n, n - v_n + u_n, n - v_n + 2u_n, \cdots, k - v_n\}.$$

Notice that, for all $t = \frac{\ell}{n}$ with $\ell \in \Pi_{n,k}$, we can find $n' > n$, $k' > k$ and $\ell'_1, \ell'_2 \in \Pi_{n',k'}$ such that $\frac{\ell'_1}{n} \leq t \leq \frac{\ell'_2}{n}$. This relation holds vice versa. It shows that the previous asymptotic results still hold by computing $\widehat{C}_{k,\ell}$ for $\ell \in \Pi^0_{n,k}$.
Moreover, if $b \equiv c > 0$ is a constant function, according to (4.10), we have

$$P\{\tau(n) < \infty\} = P\Big\{ \sup_{k>n} \max_{\ell \in \Pi^0_{n,k}} \widehat{C}_{k,\ell} > c \Big\}. \tag{4.14}$$

Thus, denote

$$\widehat{C}_k = \max_{\ell \in \Pi^0_{n,k}} \widehat{C}_{k,\ell} \quad \text{for any} \quad k > n.$$

The procedure is monitored from $k = n+1$ to $k = n+500$. The set $\{n+1, \cdots, n+500\}$ is called monitoring period. We evaluated the performance with $v_n = [\log n]$, $[(\log n)^{3/2}]$, $[(\log n)^2]$, $[(\log n)^3]$ and we recommend to use $v_n = [(\log n)^{3/2}]$ for linear model and $v_n = [(\log n)^2]$ for ARCH-type model. The nominal level used in the sequel is $\alpha = 0.05$.

4.4.1 An illustration

We consider GARCH(1,1) process : $X_t = \sigma_t \xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$. Thus, the parameter of model is $\theta_0^* = (\alpha_0^*, \alpha_1^*, \beta_1^*)$. The historical available data are $X_1, \cdots, X_{500}$ and the monitoring period is $\{501, \cdots, 1000\}$. At the nominal level $\alpha = 0.05$, the critical values of the procedure is $C_\alpha = 2.760$. The Figure 1 is a typical realization of the statistic $(\widehat{C}_k)_{500<k\leq 1000}$. We consider a scenario without change (Figure 4.1 a-)) and a scenario with change at $k^* = n + 250 = 750$ (Figure 4.1 b-)).

Figure 4.1 a-) show that, the detector $\widehat{C}_k$ is bellow under the horizontal line which represents the limit of the critical region. On Figure 4.1 b-) we can see that, before change occurs, $\widehat{C}_k$ is bellow under the horizontal line and increases with a high speed after change. Such growth over a long period indicates that something happening in the model.

4.4.2 Monitoring mean shift in times series

Let $(X_1, \cdots, X_n)$ be an (historical) observation of a process $X = (X_t)_{t\in\mathbb{Z}}$. We assume that X satisfy

$$\begin{cases} X_t = \mu_0 + \epsilon_t & \text{for} \quad 1 \leq t \leq k^* \\ X_t = \mu_1 + \epsilon_t & \text{for} \quad t > k^* \end{cases}$$

with $k^* > n$, $\mu_0 \neq \mu_1$ and (ϵ_t) a zero mean stationary time series belongs to a class $\mathcal{M}_\mathbb{Z}(f_\theta, M_\theta)$. Under H_0, $k^* = \infty$. The monitoring procedure start at $k = n + 1$ and the aim is to test mean shift over the new observation $X_{n+1}, X_{n+2}, \cdots$. This problem can be

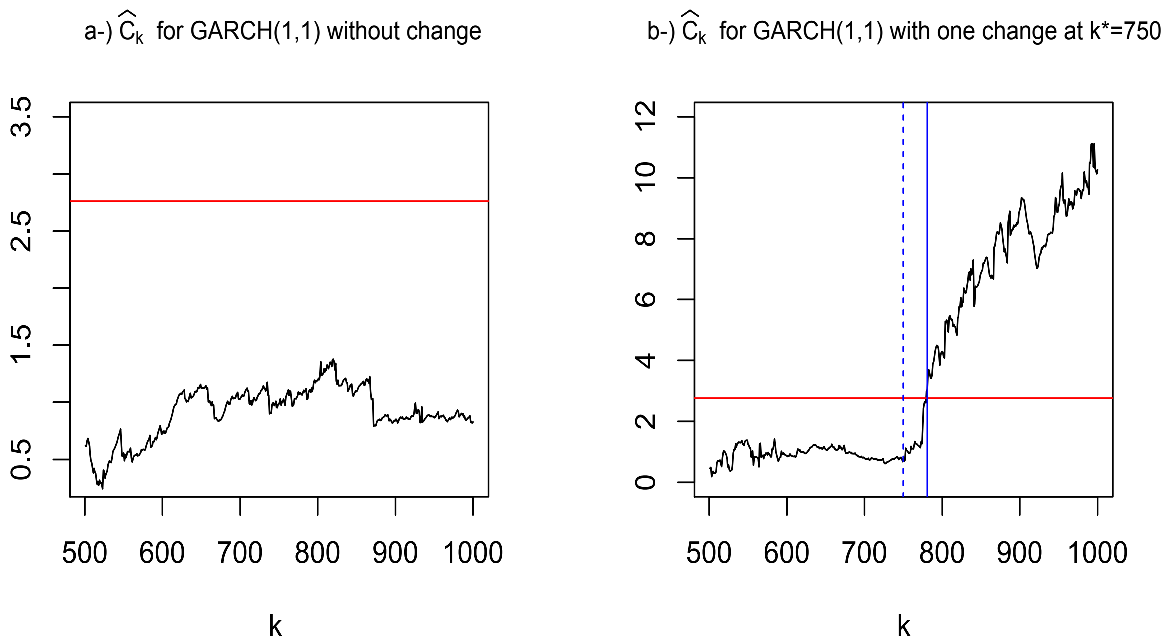

Figure 4.1 – Typical realization of the statistics $\widehat{C}_k$ with $k = 501, \cdots, 1000$. a-) The parameter $\theta_0^* = (0.01, 0.3, 0.2)$ is constant ; b-) the parameter $\theta_0^* = (0.01, 0.3, 0.2)$ changes to $\theta_1^* = (0.05, 0.5, 0.2)$ at $k^* = 750$. The horizontal solid line represents the limit of the critical region, the vertical dotted line indicates where the change occurs and the vertical solid line indicates the time where the monitoring procedure detecting the change.

see as monitoring changes in linear model (see Horváth *et al.* [34], Aue *et al.* [2]) with constant regressor. The empirical mean

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

is a consistent estimator of μ_0. The recursive residual is defined by

$$\widehat{\epsilon}_k = X_k - \overline{X}_n \ ; \ \text{for } k > n.$$

Horváth *et al.* [34] and Aue *et al.* [2] proposed the detector

$$\widehat{Q}_k = \frac{1}{\widehat{\sigma}_n}\frac{1}{c\sqrt{n}(\frac{k}{n})(1-\frac{n}{k})^{\gamma}}\Big|\sum_{i=n+1}^{k}\widehat{\epsilon}_i\Big| \quad k > n\,;\, c > 0\,;\, 0 \leq \gamma < 1/2$$

where $\widehat{\sigma}_n^2$ is an consistent estimator of the long-run variance

$$\sigma^2 = \lim_{n\to\infty}\frac{1}{n}Var(\sum_{i=1}^{n}\epsilon_i).$$

If the process (ϵ_t) are uncorrelated (for instance GARCH-type model), empirical variance of the historical data can be used as estimator of σ^2. If (ϵ_t) are correlated, the popular

Bartlett estimator (see [16]) can be used. Under some regular conditions, it hold that (see [34] and [2])

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = P\Big\{ \sup_{0<s<1} \frac{|W_1(s)|}{s^\gamma} > c\Big\}.$$

Hence, at a nominal level $\alpha = 0.05$, the critical value of the test is the $(1-\alpha)$-quantile of the distribution of $\sup\limits_{0<s<1} \|W_1(s)\|/s^\gamma$. When $\gamma = 0$, these quantiles are known (see Table 1 of [58] for values obtained through a Monte Carlo simulation).

We compare our procedure to the previous residuals CUSUM one (with $\gamma = 0$) in two situations

1. (ϵ_t) is an AR(1) process ; $\epsilon_t = \phi_1^*\epsilon_{t-1} + \xi_t$ with $\phi_1^* = 0.2$;
2. (ϵ_t) is a GARCH(1,1) process ; $\epsilon_t = \sigma_t\xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^*\epsilon_{t-1}^2 + \beta_1^*\sigma_{t-1}^2$ and $(\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.01, 0.3, 0.2)$.

The historical sample size are $n = 500$ and $n = 1000$. These procedure are evaluated at times $k = n+100, n+200, n+300, n+400, n+500$. Tables 4.2 and 4.3 indicate the empirical levels and the empirical powers based of 200 replications. The elementary statistics of the empirical detection delay are reported in Tables 4.4.

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Emp. levels	$n = 500$	$\widehat{C}_k$	0.000	0.000	0.010	0.015	0.015
		$\widehat{Q}_k$	0.000	0.005	0.005	0.010	0.015
	$n = 1000$	$\widehat{C}_k$	0.000	0.000	0.000	0.005	0.010
		$\widehat{Q}_k$	0.000	0.000	0.005	0.000	0.010
Emp. powers	$n = 500$	$\widehat{C}_k$	0.310	1	1	1	1
	$(k^* = n+50)$	$\widehat{Q}_k$	0.335	1	1	1	1
	$n = 500$	$\widehat{C}_k$	0.000	0.000	0.190	1	1
	$(k^* = n+250)$	$\widehat{Q}_k$	0.000	0.000	0.130	0.965	1
	$n = 1000$	$\widehat{C}_k$	0.075	1	1	1	1
	$(k^* = n+50)$	$\widehat{Q}_k$	0.095	1	1	1	1
	$n = 1000$	$\widehat{C}_k$	0.000	0.000	0.135	1	1
	$(k^* = n+250)$	$\widehat{Q}_k$	0.000	0.000	0.075	0.980	1

Table 4.2 – Empirical levels and powers for monitoring means shift in AR(1) with $\phi_1^* = 0.2$. The empirical levels are computed when $\mu_0 = 0$ and the empirical powers are computed when the mean $\mu_0 = 0$ changes to $\mu_1 = 1.2$.

The results of Table 4.2 and Table 4.3 show that both procedure based on detectors $\widehat{C}_k$ and $\widehat{Q}_k$ are more conservative. One can also see that, increasing length n of historical data reduces the size distortion of these procedures. This is due to the fact that the length of monitoring period is fixed and not increasing with n.
Under H_1, the change have been detected before the monitoring time $k = n+500$. But, as we mentioned above, the challenge of this problem is the detection delay. For this criteria,

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Emp. levels	$n=500$	$\widehat{C}_k$	0.005	0.015	0.030	0.055	0.060
		$\widehat{Q}_k$	0.000	0.000	0.005	0.010	0.010
	$n=1000$	$\widehat{C}_k$	0.000	0.005	0.005	0.010	0.015
		$\widehat{Q}_k$	0.000	0.000	0.000	0.010	0.010
Emp. powers	$n=500$	$\widehat{C}_k$	1	1	1	1	1
	$(k^*=n+50)$	$\widehat{Q}_k$	1	1	1	1	1
	$n=500$	$\widehat{C}_k$	0.010	0.015	1	1	1
	$(k^*=n+250)$	$\widehat{Q}_k$	0.000	0.000	0.920	1	1
	$n=1000$	$\widehat{C}_k$	0.995	1	1	1	1
	$(k^*=n+50)$	$\widehat{Q}_k$	0.985	1	1	1	1
	$n=1000$	$\widehat{C}_k$	0.000	0.000	0.980	1	1
	$(k^*=n+250)$	$\widehat{Q}_k$	0.000	0.000	0.765	1	1

Table 4.3 – Empirical levels and powers for monitoring means shift in GARCH(1,1) with $(\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.01, 0.3, 0.2)$. The empirical levels are computed when $\mu_0 = 0$ and the empirical powers are computed when the mean $\mu_0 = 0$ changes to $\mu_1 = 0.3$.

it is seen in Table 4.4 that for the mean shift in AR process, our procedure works well as Horváth *et al.'s* procedure when the change occurs at the beginning of the monitoring ($k^* = n+50$) and it is little better when the change occurs long time after the beginning of the monitoring ($k^* = n+250$). For the mean shift in GARCH process, our test procedure outperforms the Horváth *et al.'s* test in terms of mean, Q_1, median and Q_3 of detection delay.

4.4.3 Monitoring parameter change in AR(1) and GARCH(1,1) models

In this subsection, we present some simulations results for monitoring parameter change in AR(1) and GARCH(1,1) models and compare our procedure to that proposed by Na *et al.* [58]. First recall that if the boundary function $b \equiv c > 0$ is constant, this procedure is based on the relation

$$P\{\tau(n) < \infty\} = P\Big\{ \sup_{k>n} \widehat{D}_k > c \Big\}$$

where

$$\widehat{D}_k := \sqrt{n} \big\| \widehat{G}(T_{1,n})^{-1/2} \widehat{F}(T_{1,n}) \big(\widehat{\theta}(T_{1,k}) - \widehat{\theta}(T_{1,n}) \big) \big\|.$$

Na *et al.* show that under H_0,

$$\lim_{n\to\infty} P\{\tau(n) < \infty\} = \lim_{n\to\infty} P\Big\{ \sup_{k>n} \widehat{D}_k > c \Big\} = P\Big\{ \sup_{0<s<1} \|W_d(s)\| > c \Big\}.$$

Hence, at a nominal level α, the critical value of their procedure is the $(1-\alpha)$-quantile of the distribution of $\sup_{0<s<1} \|W_d(s)\|$ which can be found in Table 1 of [58].

The comparisons are made in the followings situations.

$\widehat{d}_n$			Mean	SD	Min	Q_1	Med	Q_3	Max
AR(1)	$n = 500\,;\ k^* = n + 50$	$\widehat{C}_k$	54.74	14.95	18	44	54	64	103
		$\widehat{Q}_k$	53.78	14.72	16	43	54	63	102
	$n = 500\,;\ k^* = n + 250$	$\widehat{C}_k$	63.14	23.18	12	45	61	77	135
		$\widehat{Q}_k$	72.70	21.47	7	56	71.5	90	139
	$n = 1000\,;\ k^* = n + 50$	$\widehat{C}_k$	75.84	14.19	37	66	75	83	114
		$\widehat{Q}_k$	72.60	13.23	41	63	73	82	111
	$n = 1000\,;\ k^* = n + 250$	$\widehat{C}_k$	76.24	19.15	23	60	76	89	140
		$\widehat{Q}_k$	86.82	22.57	27	70	85	100	151
GARCH(1,1)	$n = 500\,;\ k^* = n + 50$	$\widehat{C}_k$	20.21	6.15	1	16	20	24	35
		$\widehat{Q}_k$	27.06	4.52	16	24	27	30	44
	$n = 500\,;\ k^* = n + 250$	$\widehat{C}_k$	25.53	8.04	3	20	25	31	50
		$\widehat{Q}_k$	35.40	10.01	13	28	35	41	62
	$n = 1000\,;\ k^* = n + 50$	$\widehat{C}_k$	28.43	7.41	6	24	28	33	51
		$\widehat{Q}_k$	36.98	5.09	21	33	37	40	48
	$n = 1000\,;\ k^* = n + 250$	$\widehat{C}_k$	31.16	8.52	4	26	33	39	53
		$\widehat{Q}_k$	44.35	10.04	14	37	45	50	71

Table 4.4 – Elementary statistics of the empirical detection delay for monitoring mean shift in AR(1) and GARCH(1,1).

1. For **AR(1) model** : $X_t = \phi_1^* X_{t-1} + \xi_t$. Under H_0, $\theta_0 = \phi_1^* = 0.2$; and $\theta_0 = 0.2$ changes to $\theta_1 = -0.5$ under H_1.
2. For **GARCH(1,1) model** : $X_t = \sigma_t \xi_t$ with $\sigma_t^2 = \alpha_0^* + \alpha_1^* X_{t-1}^2 + \beta_1^* \sigma_{t-1}^2$. Under H_0, $\theta_0 = (\alpha_0^*, \alpha_1^*, \beta_1^*) = (0.05, 0.3, 0.2)$; and $\theta_0 = (0.01, 0.3, 0.2)$ changes to $\theta_1 = (0.05, 0.5, 0.2)$ under H_1.

The historical sample size are $n = 500$ and $n = 1000$. These procedure are evaluated at times $k = n+100, n+200, n+300, n+400, n+500$. Tables 4.5 and 4.6 indicate the empirical levels and the empirical powers based of 200 replications. The elementary statistics of the empirical detection delay are reported in Tables 4.7.

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Emp. levels	$n = 500$	$\widehat{C}_k$	0.000	0.000	0.010	0.010	0.035
		$\widehat{D}_k$	0.000	0.000	0.000	0.000	0.025
	$n = 1000$	$\widehat{C}_k$	0.000	0.000	0.000	0.010	0.025
		$\widehat{D}_k$	0.000	0.000	0.000	0.000	0.020
Emp. powers	$n = 500$	$\widehat{C}_k$	0.335	1	1	1	1
	$(k^* = n + 50)$	$\widehat{D}_k$	0.175	0.985	1	1	1
	$n = 500$	$\widehat{C}_k$	0.000	0.000	0.180	990	1
	$(k^* = n + 250)$	$\widehat{D}_k$	0.000	0.000	0.095	0.865	1
	$n = 1000$	$\widehat{C}_k$	0.065	0.995	1	1	1
	$(k^* = n + 50)$	$\widehat{D}_k$	0.090	0.975	1	1	1
	$n = 1000$	$\widehat{C}_k$	0.000	0.000	0.140	0.990	1
	$(k^* = n + 250)$	$\widehat{D}_k$	0.000	0.000	0.075	0.855	0.995

TABLE 4.5 – Empirical levels and powers for monitoring parameter change in AR(1) process. The empirical levels are computed when $\theta_0 = \phi_1^* = 0.2$ is constant and the empirical powers are computed when $\theta_0 = 0.2$ changes to $\theta_1 = -0.5$.

The processes AR and GARCH considered have zero mean. Contrary to the mean shift studied above, this mean is not estimated. For AR model, it appears in Table 4.5 that both procedure based on detector $\widehat{C}_k$ and $\widehat{D}_k$ are conservative. This is not the case for GARCH model (Table 4.6) . The high size distortions when $n = 500$ is due to the difficulty to estimate the parameter of GARCH model. This size distortion decreases when n increases and Corollary 4.3.1 enures that with infinite monitoring period, the empirical level tends to the nominal one as $n \to \infty$.

For both cases AR and GARCH model, the procedure based on detector $\widehat{C}_k$ detect the change after the monitoring time $k = n + 500$. Unlike Na *et al.* [58], we consider a scenario of GARCH model with moderate change in parameter, it is seen in Table 4.6 that the paocedure based on detector $\widehat{D}_k$ provides unsatisfactory results. At the monitoring time $k = n + 500$, it is not sure that the change must be detected. This is not suppressing according to the comment of subsection 4.3.1.

Table 4.7 indicates the distribution of the detection delay $\widehat{d}_n$. We can see in Table 4.7 (even in Table 4.4) that for our procedure, the relation $\widehat{d}_{1000} \leq \sqrt{2}\ \widehat{d}_{500}$ is satisfied. This

		k	$n+100$	$n+200$	$n+300$	$n+400$	$n+500$
Emp. levels	$n=500$	$\widehat{C}_k$	0.010	0.025	0.040	0.095	0.105
		$\widehat{D}_k$	0.010	0.015	0.040	0.040	0.055
	$n=1000$	$\widehat{C}_k$	0.000	0.000	0.030	0.045	0.055
		$\widehat{D}_k$	0.000	0.000	0.010	0.015	0.035
Emp. powers	$n=500$	$\widehat{C}_k$	0.890	1	1	1	1
	$(k^*=n+50)$	$\widehat{D}_k$	0.390	0.855	0.930	0.965	0.985
	$n=500$	$\widehat{C}_k$	0.010	0.030	0.825	1	1
	$(k^*=n+250)$	$\widehat{D}_k$	0.010	0.020	0.270	0.805	0.915
	$n=1000$	$\widehat{C}_k$	0.835	1	1	1	1
	$(k^*=n+50)$	$\widehat{D}_k$	0.310	0.970	0.990	0.995	0.995
	$n=1000$	$\widehat{C}_k$	0.000	0.005	0.685	1	1
	$(k^*=n+250)$	$\widehat{D}_k$	0.000	0.000	0.250	0.955	0.990

Table 4.6 – Empirical levels and powers for monitoring parameter change in GARCH(1,1) process. The empirical levels are computed when $\theta_0=(\alpha_0^*,\alpha_1^*,\beta_1^*)=(0.01,0.3,0.2)$ is constant and the empirical powers are computed when $\theta_0=(0.01,0.3,0.2)$ changes to $\theta_1=(0.05,0.5,0.2)$.

confirm the result of Theorem 4.3.2 that the detection delay is bounded by $\sqrt{n}$. It is also seen that, mean, Q_1, median and Q_3 of our test are shorter than Na *et al.*'s one. The results of Table 4.5, 4.6 and 4.7 show that, our test is uniformly better and the procedure based on detector $\widehat{C}_k$ is highly recommended.

4.5 Proofs of the main results

Let us prove first some useful lemmas. In the sequel, for any $x\in\mathbb{R}$, $[x]$ denotes the integer part of x. Let $(\psi_n)_n$ and $(r_n)_n$ be sequences of random variables. Throughout this section, we use the notation $\psi_n=o_P(r_n)$ to mean : for all $\varepsilon>0$, $P(|\psi_n|\geq\varepsilon|r_n|)\to 0$ as $n\to\infty$. Write $\psi_n=O_P(r_n)$ to mean : for all $\varepsilon>0$, there exists $C>0$ such that $P(|\psi_n|\geq C|r_n|)\leq\varepsilon$ for n large enough.

Recall that the historical available observations $X_1,\cdots,X_n$ is a trajectory of a process $\mathcal{M}_{\mathbb{Z}}(M_{\theta_0^*},f_{\theta_0^*})$.
Let $k\geq n\geq 2$ and $T_{1,n}=\{1,\cdots,n\}$, $T_{\ell,k}=\{\ell,\ell+1,\cdots,k\}$ with $\ell\in\Pi_{n,k}=\{v_n,v_n+1,\cdots,k-v_n\}$, and define

$$C_{k,\ell}:=\sqrt{n}\,\frac{k-\ell}{k}\big\|G^{-1/2}F\cdot\big(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n})\big)\big\|,$$

with $\widehat{\theta}$ defined in (4.6).

Lemma 4.5.1. Under assumptions of Theorem 4.3.1,

$$\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\big|\widehat{C}_{k,\ell}-C_{k,\ell}\big|=o_P(1)\ \ \text{as } n\to\infty.$$

$\widehat{d}_n$			Mean	SD	Min	Q_1	Med	Q_3	Max
AR(1)	$n = 500\,;\ k^* = n + 50$	$\widehat{C}_k$	55.36	18.75	9	42	56	67	121
		$\widehat{D}_k$	71.54	38.44	2	52.75	69	89	167
	$n = 500\,;\ k^* = n + 250$	$\widehat{C}_k$	66.81	25.27	5	49	65	83	149
		$\widehat{D}_k$	97.80	39.42	21	68	89	123	222
	$n = 1000\,;\ k^* = n + 50$	$\widehat{C}_k$	75.13	19.87	24	62	74	90	147
		$\widehat{D}_k$	87.70	28.72	14	66	85	109	195
	$n = 1000\,;\ k^* = n + 250$	$\widehat{C}_k$	76.89	26.16	15	56	77	96	172
		$\widehat{D}_k$	101.20	37.97	20	75	96	129	245
GARCH(1,1)	$n = 500\,;\ k^* = n + 50$	$\widehat{C}_k$	29.41	15.84	4	22	31	40	98
		$\widehat{D}_k$	86.05	90.50	2	36	61	99	416
	$n = 500\,;\ k^* = n + 250$	$\widehat{C}_k$	38.02	19.33	5	27	37	44	113
		$\widehat{D}_k$	87.72	50.96	1	49.25	79	112	236
	$n = 1000\,;\ k^* = n + 50$	$\widehat{C}_k$	41.96	13.93	3	32	41	48	94
		$\widehat{D}_k$	71.29	37.12	6	46	66	88	287
	$n = 1000\,;\ k^* = n + 250$	$\widehat{C}_k$	44.99	17.16	5	35	41	52	117
		$\widehat{D}_k$	75.78	35.10	7	52	71	95	198

Table 4.7 – Elementary statistics of the empirical detection delay for monitoring parameter change in AR(1) and GARCH(1,1).

Proof. For any $n \geq 1$, we have

$$\sup_{k>n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} |\widehat{C}_{k,\ell} - C_{k,\ell}| = \frac{1}{\inf_{s>0} b(s)} \sup_{k>n} \max_{\ell \in \Pi_{n,k}} |\widehat{C}_{k,\ell} - C_{k,\ell}|.$$

Now, proceed similarly as in the proof of Lemma 5.3 of [40].

Lemma 4.5.2. Under assumptions of Theorem 4.3.1

$$\sup_{k>n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \|(k-\ell) F \cdot (\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})) - 2(\frac{\partial}{\partial\theta} L(T_{\ell,k}, \theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n}, \theta_0^*))\| \\ = o_P(1) \ \text{ as } n \to \infty.$$

Proof. Let $k \geq n$ and $T \subset \{1, \cdots, k\}$. By applying the Taylor expansion to the coordinates of $\partial \widehat{L}(T, \cdot)/\partial\theta$, and using the fact that $\partial \widehat{L}(T, \widehat{\theta}(T))/\partial\theta = 0$ we have

$$\frac{2}{\mathrm{Card}(T)} \frac{\partial}{\partial\theta} \widehat{L}(T, \theta_0^*) = \widetilde{F}(T) \cdot (\widehat{\theta}(T) - \theta_0^*) \ \text{ where } \ \widetilde{F}(T) = -2\Big(\frac{1}{\mathrm{Card}(T)} \frac{\partial^2 \widehat{L}(T, \widetilde{\theta}_i(T))}{\partial\theta\partial\theta_i}\Big)_{1\leq i \leq d}$$

for some $\widetilde{\theta}_i(T)$ between $\widehat{\theta}(T)$ and θ_0^*.
Hence for any $\ell \in \Pi_{n,k}$

$$F \cdot (\widehat{\theta}(T_{\ell,k}) - \theta_0^*) = \frac{2}{k-\ell} \frac{\partial}{\partial\theta} L(T_{\ell,k}, \theta_0^*) + \big(F - \widetilde{F}(T_{\ell,k})\big)\big(\widehat{\theta}(T_{\ell,k}) - \theta_0^*\big) \\ + \frac{2}{k-\ell}\Big(\frac{\partial}{\partial\theta} \widehat{L}(T_{\ell,k}, \theta_0^*) - \frac{\partial}{\partial\theta} L(T_{\ell,k}, \theta_0^*)\Big).$$

and

$$F\,(\widehat{\theta}(T_{1,n}) - \theta_0^*) = \frac{2}{n} \frac{\partial}{\partial\theta} L(T_{1,n}, \theta_0^*) + \big(F - \widetilde{F}(T_{1,n})\big)\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) + \frac{2}{n}\Big(\frac{\partial}{\partial\theta} \widehat{L}(T_{1,n}, \theta_0^*) - \frac{\partial}{\partial\theta} L(T_{1,n}, \theta_0^*)\Big).$$

Therefore, for any $\ell \in \Pi_{n,k}$

$$\begin{aligned} &\frac{\sqrt{n}}{k}\Big((k-\ell) F\,(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})) - 2\Big(\frac{\partial}{\partial\theta} L(T_{\ell,k}, \theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n}, \theta_0^*)\Big)\Big) \\ &\quad = \sqrt{n}\frac{k-\ell}{k}\big(F - \widetilde{F}(T_{\ell,k})\big)\big(\widehat{\theta}(T_{\ell,k}) - \theta_0^*\big) + 2\frac{\sqrt{n}}{k}\Big(\frac{\partial}{\partial\theta} \widehat{L}(T_{\ell,k}, \theta_0^*) - \frac{\partial}{\partial\theta} L(T_{\ell,k}, \theta_0^*)\Big) \\ &\quad - \sqrt{n}\frac{k-\ell}{k}\big(F - \widetilde{F}(T_{1,n})\big)\big(\widehat{\theta}(T_{1,n}) - \theta_0^*\big) - 2\frac{k-\ell}{k}\frac{1}{\sqrt{n}}\Big(\frac{\partial}{\partial\theta} \widehat{L}(T_{1,n}, \theta_0^*) - \frac{\partial}{\partial\theta} L(T_{1,n}, \theta_0^*)\Big). \end{aligned} \tag{4.15}$$

For $k > n$ and with some $\ell_k \in \Pi_{n,k}$, we have

$$\max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|F - \widetilde{F}(T_{\ell,k})\|\,\|\widehat{\theta}(T_{\ell,k}) - \theta_0^*\| \leq \frac{1}{\underset{s>0}{\mathrm{Inf}} b(s)} \sqrt{k-\ell_k} \|F - \widetilde{F}(T_{\ell_k,k})\|\,\|\widehat{\theta}(T_{\ell_k,k}) - \theta_0^*\|.$$

According to [8] and [9], $\|F - \widetilde{F}(T_{\ell_k,k})\| = o_P(1)$ and $\|\widehat{\theta}(T_{\ell_k,k}) - \theta_0^*\| = O_P(1/\sqrt{k-\ell_k})$ as $k - \ell_k \to \infty$. Hence

$$\sup_{k>n} \max_{\ell \in \Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|F - \widetilde{F}(T_{\ell,k})\|\,\|\widehat{\theta}(T_{\ell,k}) - \theta_0^*\| = o_P(1) \ \text{ as } \ n \to \infty. \tag{4.16}$$

Similar arguments imply that

$$\sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|F - \widetilde{F}(T_{1,n})\| \, \|\widehat{\theta}(T_{1,n}) - \theta_0^*\| = o_P(1) \quad \text{as} \quad n \to \infty. \quad (4.17)$$

For $k > n$ and for some $\ell_k \in \Pi_{n,k}$, we have

$$\max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k}, \theta_0^*) - \frac{\partial}{\partial\theta}L(T_{\ell,k}, \theta_0^*)\|$$
$$\leq \frac{1}{\underset{s>0}{\operatorname{Inf}} b(s)} \frac{1}{\sqrt{k-\ell_k}} \|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell_k,k}, \theta_0^*) - \frac{\partial}{\partial\theta}L(T_{\ell_k,k}, \theta_0^*)\|.$$

According to [8], $\dfrac{1}{\sqrt{k-\ell_k}}\|\dfrac{\partial}{\partial\theta}\widehat{L}(T_{\ell_k,k}, \cdot) - \dfrac{\partial}{\partial\theta}L(T_{\ell_k,k}, \cdot)\|_\Theta = o_P(1)$ as $k - \ell_k \to \infty$. Hence

$$\sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \|\frac{\partial}{\partial\theta}\widehat{L}(T_{\ell,k}, \theta_0^*) - \frac{\partial}{\partial\theta}L(T_{\ell,k}, \theta_0^*)\| = o_P(1) \quad \text{as} \quad n \to \infty. \quad (4.18)$$

Similar arguments show that

$$\sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{k-\ell}{k} \frac{1}{\sqrt{n}} \|\frac{\partial}{\partial\theta}\widehat{L}(T_{1,n}, \theta_0^*) - \frac{\partial}{\partial\theta}L(T_{1,n}, \theta_0^*)\| = o_P(1) \quad \text{as} \quad n \to \infty. \quad (4.19)$$

Thus, Lemma 4.5.2 follows from (4.15), (4.16), (4.17), (4.18) and (4.19).

Lemma 4.5.3. Under assumptions of Theorem 4.3.1

$$\sup_{k>n} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \sqrt{n}\frac{k-\ell}{k} \|F\cdot\big(\widehat{\theta}(T_{\ell,k}) - \widehat{\theta}(T_{1,n})\big)\| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{t>1} \sup_{0<s<t} \frac{\|W_G(s) - sW_G(1)\|}{t\, b(s)}$$

where W_G is a d-dimensional Gaussian centered process with covariance matrix $\mathbb{E}(W_G(s)W_G(\tau)') = \min(s,\tau)G$.

Proof. For $k > n$ and $\ell \in \Pi_{n,k}$, we have

$$2\frac{\sqrt{n}}{k}\left(\frac{\partial}{\partial\theta}L(T_{\ell,k}, \theta_0^*) - \frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n}, \theta_0^*)\right) = -\frac{n}{k}\frac{1}{\sqrt{n}}\left(\sum_{i=\ell+1}^{k} \frac{\partial q_i(\theta_0^*)}{\partial\theta} - \frac{k-\ell}{n}\sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\right).$$

Now we are going to proceed in two steps.
Step 1. Let $T > 1$. We have

$$\max_{n<k<nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{\partial}{\partial\theta}L(T_{\ell,k}, \theta_0^*) - \frac{k-\ell}{n}\frac{\partial}{\partial\theta}L(T_{1,n}, \theta_0^*)\|$$
$$= \max_{n<k<nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \|\sum_{i=\ell}^{k} \frac{\partial q_i(\theta_0^*)}{\partial\theta} - \frac{k-\ell}{n}\sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\|$$
$$= \max_{t\in\{1,1+\frac{1}{n},\cdots,T\}} \max_{s\in\{1-\frac{v_n}{n},2-\frac{v_n}{n},\cdots,t-\frac{v_n}{n}\}} \frac{1}{b(\frac{[nt]-[ns]}{n})} \frac{n}{[nt]} \|\frac{1}{\sqrt{n}}\left(\sum_{i=[ns]}^{[nt]} \frac{\partial q_i(\theta_0^*)}{\partial\theta} - \frac{[nt]-[ns]}{n}\sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\right)\|.$$

Define the set $S := \{(t,s) \in [1,T] \times [1,T] /\ s < t\}$. According to [8], $\big(\dfrac{\partial q_i(\theta_0^*)}{\partial \theta}\big)_{t\in\mathbb{Z}}$ is a stationary ergodic martingale difference sequence with covariance matrix G. By Cramér-Wold device (see [18] p. 206), it holds that

$$\frac{1}{\sqrt{n}} \sum_{i=[ns]+1}^{[nt]} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \ \xrightarrow[n\to\infty]{\mathcal{D}(S)} \ W_G(t-s).$$

with $\xrightarrow[n\to\infty]{\mathcal{D}(S)}$ means the weak convergence on the Skorohod space $\mathcal{D}(S)$. Hence

$$\frac{1}{\sqrt{n}} \Big(\sum_{i=[ns]+1}^{[nt]} \frac{\partial q_i(\theta_0^*)}{\partial \theta} - \frac{[nt]-[ns]}{n} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial \theta} \Big) \xrightarrow[n\to\infty]{\mathcal{D}(S)} \ W_G(t-s) - (t-s) W_G(1).$$

Therefore

$$\begin{aligned}
\max_{n<k<nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \Big\| \frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*) \Big\| & \\
\xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1<t<T} \sup_{1<s<t} \frac{\|W_G(t-s) - (t-s) W_G(1)\|}{t\ b(t-s)} & \\
\xrightarrow[n\to\infty]{\mathcal{D}} \sup_{1<t<T} \sup_{1<s<t} \frac{\|W_G(s) - s\, W_G(1)\|}{t\ b(s)}. & \qquad (4.20)
\end{aligned}$$

Step 2. We will show that the limit distribution (as $n, T \to \infty$) of

$$\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \Big\| \frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*) \Big\|$$

exists and is equal to the limit distribution (as $T \to \infty$) of

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - s\, W_G(1)\|}{t\ b(s)}.$$

Let $k > nT$. We have

$$\max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{\sqrt{n}}{k} \Big\| \frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) \Big\| \leq \frac{1}{\operatorname*{Inf}_{s>0} b(s)} \frac{\sqrt{n}}{k} \Big\| \sum_{i=\ell_k+1}^{k} \frac{\partial q_i(\theta_0^*)}{\partial\theta} \Big\| \ \text{ for some } \ell_k \in \Pi_{n,k}.$$

It comes from the Hájek-Rényi-Chow inequality (see [22]) that, for any $\varepsilon > 0$

$$\lim_{T\to\infty} \lim_{n\to\infty} P\Big(\sup_{k>nT} \frac{\sqrt{n}}{k} \Big\| \sum_{i=\ell_k+1}^{k} \frac{\partial q_i(\theta_0^*)}{\partial\theta} \Big\| > \varepsilon \Big) = 0.$$

Hence

$$\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \Big\| \frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) \Big\| = o_P(1) \ \text{ as } \ T, n \to \infty. \qquad (4.21)$$

Moreover, since the function $b(\cdot)$ is non-increasing, for any $n, T > 1$, we have :

$$\begin{aligned}
\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*)\| &= \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\| \\
&\quad \times \sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{k-\ell}{k} \\
&= \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\| \\
&\quad \times \sup_{k>nT} \frac{1}{b((k-v_n)/n)} \frac{k-v_n}{k} \\
&= \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial q_i(\theta_0^*)}{\partial\theta}\| \\
&\xrightarrow[n\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|, \qquad (4.22)
\end{aligned}$$

using again the Cramèr-Wold device. It comes from (4.21) and (4.22) that

$$\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*)\| \xrightarrow[T,n\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{4.23}$$

Furthermore, since the coordinates of W_G are Brownian motions, by the law of the iterated logarithm there exists $t_0 > \exp(1)$ such as

$$s > t_0 \Rightarrow \|W_G(s)\| \leq \sqrt{s}\log(s) \ \text{ almost surely.}$$

Thus, for any $t > t_0$, we obtain almost surely

$$\sup_{1<s<t} \|W_G(s)\| \leq \sup_{1<s<t_0} \|W_G(s)\| + \sqrt{t}\log(t).$$

Therefore, for T large enough, we have

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s)\|}{t\ b(s)} \leq \frac{1}{\inf_{s>0} b(s)} \big(\frac{1}{T} \sup_{1<s<t_0} \|W_G(s)\| + \sup_{t>T} \frac{\log(t)}{\sqrt{t}}\big) \xrightarrow[T\to\infty]{\text{a.s.}} 0. \tag{4.24}$$

Finally, since $b(\cdot)$ is non-increasing, for any $T > 1$, we have

$$\sup_{t>T} \sup_{1<s<t} \frac{\|sW_G(1)\|}{t\ b(s)} = \|W_G(1)\| \sup_{t>T} \frac{1}{t} \sup_{1<s<t} \frac{s}{b(t)} = \|W_G(1)\| \sup_{t>T} \frac{1}{b(t)} = \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{4.25}$$

It comes from (4.24) and (4.25) that the limit of (4.20) satisfies when $T \to \infty$,

$$\sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - sW_G(1)\|}{t\ b(s)} \xrightarrow[T\to\infty]{\mathcal{D}} \frac{1}{\operatorname{Inf}_{s>0} b(s)} \|W_G(1)\|. \tag{4.26}$$

From **Step 1** and **Step 2** (the relations (4.20), (4.23) and (4.26)), it comes that

$$\sup_{k>nT} \max_{\ell\in\Pi_{n,k}} \frac{1}{b((k-\ell)/n)} \frac{2\sqrt{n}}{k} \|\frac{\partial}{\partial\theta} L(T_{\ell,k},\theta_0^*) - \frac{k-\ell}{n} \frac{\partial}{\partial\theta} L(T_{1,n},\theta_0^*)\| \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{t>T} \sup_{1<s<t} \frac{\|W_G(s) - sW_G(1)\|}{t\ b(s)}.$$

Hence, Lemma 4.5.3 follows from Lemma 4.5.2.

Proof of Theorem 4.3.1
We know that

$$\begin{aligned} P\{\tau(n)<\infty\} &= P\Big\{\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{\widehat{C}_{k,\ell}}{b((k-\ell)/n)}>1\Big\} \\ &= P\Big\{\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n}))\|>1\Big\}. \end{aligned}$$

Since $\widehat{G}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} G$ and $\widehat{F}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} F$, it comes from Lemma 4.5.1 and 4.5.3 that

$$\begin{aligned} &\sup_{k>n}\max_{\ell\in\Pi_{n,k}}\frac{1}{b((k-\ell)/n)}\sqrt{n}\frac{k-\ell}{k}\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot(\widehat{\theta}(T_{\ell,k})-\widehat{\theta}(T_{1,n}))\| \\ &\qquad\qquad \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{t>1}\sup_{1<s<t}\frac{\|G^{-1/2}(W_G(s)-sW_G(1))\|}{t\;b(s)}. \end{aligned}$$

Since the covariance matrix of $\{W_G(s)\ ;\ s\geq 0\}$, is $\min(s,\tau)G$, the covariance matrix of $\{G^{-1/2}W_G(s)\ ;\ s\geq 0\}$ is $\min(s,\tau)I_d$ (where I_d is the d-dimensional identity matrix). Hence Theorem 4.3.1 follows. $\square$

Proof of Corollary 4.3.1
Since $b\equiv c$ a positive constant, it follows immediately from Theorem 4.3.1 that

$$\lim_{n\to\infty} P\{\tau(n)<\infty\} = P\Big\{\sup_{t>1}\sup_{1<s<t}\frac{1}{t}\|W_d(s)-sW_d(1))\|>c\Big\}.$$

Now, it suffices to show that $\sup_{t>1}\sup_{1<s<t}\frac{1}{t}\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} U_d$.
For any $t>1$, we have

$$\sup_{1<s<t}\frac{1}{t}\,\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{1<s<t}\frac{s}{t}\,\|W_d(\frac{s-1}{s})\| = \sup_{0<u<1-1/t}\frac{1}{t(1-u)}\,\|W_d(u)\|.$$

Thus

$$\sup_{t>1}\sup_{1<s<t}\frac{1}{t}\,\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{t>1}\sup_{0<u<1-1/t}\frac{1}{t(1-u)}\,\|W_d(u)\| = \sup_{0<v<1}\sup_{0<u<v}\frac{1-v}{1-u}\,\|W_d(u)\|.$$

But, $\|W_d(u)\| \stackrel{\mathcal{D}}{=} v^{1/2}\,\|W_d(\frac{u}{v})\|$. Therefore with $u=u'v$,

$$\sup_{0<v<1}\sup_{0<u<v}\frac{1-v}{1-u}\,\|W_d(u)\| = \sup_{0<v<1}\sup_{0<u'<1}\frac{(1-v)v^{1/2}}{1-u'v}\,\|W_d(u')\|.$$

It remains to compute $\sup_{0<v<1}\frac{(1-v)v^{1/2}}{1-u'v}$. Classical computations show that this supremum is obtained by $v=2\left(3-u'+\sqrt{(9-u')(1-u')}\right)^{-1}$ and therefore

$$\begin{aligned} &\sup_{t>1}\sup_{1<s<t}\frac{1}{t}\,\|W_d(s)-sW_d(1))\| \stackrel{\mathcal{D}}{=} \sup_{0<u'<1} f(u')\,\|W_d(u')\| \\ &\qquad \text{with}\quad f(u')=\frac{\sqrt{9-u'}+\sqrt{1-u'}}{\sqrt{9-u'}+3\sqrt{1-u'}}\Big(\frac{2}{3-u'+\sqrt{(9-u')(1-u')}}\Big)^{1/2}. \end{aligned} \tag{4.27}$$

Hence,

$$\sup_{t>1} \sup_{1<s<t} \frac{1}{t}\|W_d(s) - sW_d(1))\| \stackrel{\mathcal{D}}{=} U_d. \ \square$$

Proof of Theorem 4.3.2
Denote $k_n = k^* + n^\delta$ for $\delta \in (1/2, 1)$. For n large enough, we have $v_n < n^\delta$ and thus $k_n - v_n = k^* + n^\delta - v_n \geq k^*$. Moreover, since $k^* > n$ then $k^* \in \Pi_{n,k}$ for n large enough. In addition, since $k^* = k^*(n) \geq n$ and $\limsup_{n\to\infty} k^*(n)/n < \infty$, there exists $c_0 > 1$ such that $k^* \leq c_0 n$ for n large enough. Hence, according to assumption **B**, there exists a constant $c > 0$ such that

$$\begin{aligned}
\max_{\ell\in\Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} &= \max_{\ell\in\Pi_{n,k_n}} \frac{1}{b((k_n-\ell)/n)}\sqrt{n}\frac{k_n-\ell}{k_n}\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot\big(\widehat{\theta}_{k_n}(T_{\ell,k_n}) - \widehat{\theta}(T_{1,n})\big) \\
&\geq \frac{1}{b((k_n-k^*)/n)}\sqrt{n}\frac{k_n-k^*}{k_n}\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\cdot\big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big) \\
&\geq c\,\sqrt{n}\ \frac{n^\delta}{k^*+n^\delta}\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\| \\
&\geq c\ \frac{n^{1/2+\delta}}{c_0 n+n^\delta}\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\| \\
&\geq c\ \frac{n^{\delta-1/2}}{(c_0+1)}\big\|\widehat{G}(T_{1,n})^{-1/2}\widehat{F}(T_{1,n})\big(\widehat{\theta}_{k_n}(T_{k^*,k_n}) - \widehat{\theta}(T_{1,n})\big)\big\|. \qquad (4.28)
\end{aligned}$$

According to [8] and [40], $\widehat{G}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} G$, $\widehat{F}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} F$, $\widehat{\theta}(T_{1,n}) \xrightarrow[n\to\infty]{a.s.} \theta_0^*$ and $\widehat{\theta}_{k_n}(T_{k^*,k_n})) \xrightarrow[n\to\infty]{a.s.} \theta_1^*$. Since G is symmetric positive definite, F is invertible, $\theta_0^* \neq \theta_1^*$ and $\delta > 1/2$, then (4.28) implies that

$$\max_{\ell\in\Pi_{n,k_n}} \frac{\widehat{C}_{k_n,\ell}}{b((k_n-\ell)/n)} \xrightarrow[n\to\infty]{a.s.} \infty. \ \square$$

Chapitre 5

Application aux débits du bassin versant de la Sanaga

Nous avons dévéloppé deux procédures off-line de détection de rupture dans les processus causaux. La première procédure consiste à estimer directement les éventuels instants de rupture en utilisant un contraste pénalisé et la deuxième procédure est basée sur un test successif en utilisant un algorithme de type ISCC (Iterated Cumulative Sums of Squares). Nous allons maintenant appliquer ces procédures aux débits du bassin versant de la Sanaga.

5.1 Présentation du bassin versant de la Sanaga et des données

5.1.1 Le bassin versant de la Sanaga

La Sanaga est le plus grand fleuve du Cameroun avec un bassin versant de 133000 Km^2 ; soit plus de $\frac{1}{4}$ de la superficie totale du Cameroun. Ce bassin qui s'étend entre les parallèles 3°30 et 7°30 dispose d'un potentiel énergetique important. Près de 90% de l'énergie hydroelectrique (soit plus de 70% de l'énergie électrique) est produite sur la Sanaga à travers les usines hydroelectriques de Song-Loulou et d'Edéa.

FIGURE 5.1 – Deux vues de la Sanaga au niveau des barrages hydroelectriques d'Edéa (gauche) et de Song-Loulou (droite).

Le régime naturel de la Sanaga présente de très grandes variations ; ceci a motivé la construction de trois barrages de retenu (Mbakaou 1970, Bamendjing 1974 et Magba 1987) afin de mieux réguler les débits de ce fleuve. Malgré cela, le Cameroun subit ces dernières années une crise énergetique sérieuse dont la forme la plus visible est le délestage régulier. Les besoins en énergie électrique ont une croissance exponentielle, ainsi la quantité d'énergie produite (notamment sur la Sanaga) devient inssuffisante. C'est ainsi que le pays s'est lancé dans la construction de nouvelles infrastructures de production d'énergie (le barrage hydroélectrique de Lom Pangar...). En plus du fait que la quantité d'énergie hydroélectrique produite sur la Sanaga devient de plus en plus insuffisante, le débit de ce fleuve présente des variations irrégulières, ceci s'est manifesté par des délestages plus accrus durant certaines années. Nous allons dans cette partie éssayer de donner une signification statistique à ces variations irrégulières.

5.1.2 Présentation des données

Les données que nous disposons proviennent d'AES-Sonel (la société nationale d'électricité du Cameroun). Se sont les débits (en m^3/s) naturels journaliers reconstitués du bassin versant intermédiaire de la Sanaga à la station de Sogmbengué. Ces débits sont calculés du 1er juillet 1988 au 30 juin 2005. Sogmbengué est le dernier point de contrôle avant les ouvrages de production d'énergie hydroélectrique. C'est donc à cette station qu'on peut extraire le maximum d'information sur le débit du fleuve au niveau des organes de production d'énergie. La Figure 5.2 présente l'évolution de ces débits journaliers reconstitués.

La Figure 5.2 ne permet pas de dégager une tendance particulière, en revanche elle montre une forte variation saisonnière que nous allons étudier dans la suite.

5.1.3 Etude des variations saisonnières

Les séries hydrologiques observées à une fréquence inferieure à un an présentent généralement de très fortes variations saisonnières. C'est la cas des séries journalières, mensuelles,... Cette saisonnalité est principalement due aux variations saisonnières du climat. Pour mieux observer les variations saisonnières, on range la série sous la forme matricielle

$$\mathbb{X} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,365} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,365} \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ x_{N,1} & x_{N,2} & \cdots & x_{N,365} \end{pmatrix}$$

où N est le nombre d'années et $x_{i,j}$ représente la valeur de la série observée le jour i de l'année j. Pour un jour i, on définit la moyenne et l'écart-type empirique par

$$\overline{x}_i = \frac{1}{N}\sum_{j=1}^{N} x_{j,i} \quad ; \quad \widehat{\sigma}_i = \Big(\frac{1}{N}\sum_{j=1}^{N}(x_{j,i} - \overline{x}_i)^2\Big)^{1/2}.$$

La Figure 5.3 donne l'évolution de la moyenne et de l'écart-type empirique au cours de l'année.

La Figure 5.3 montre une forte variation saisonnière dans les débits. Les débits sont élevés durant les mois de septembre et octobre et les mois de février, mars et avril correspondent à la période où les débits sont plus faibles. Ces variations étant régulières (donc

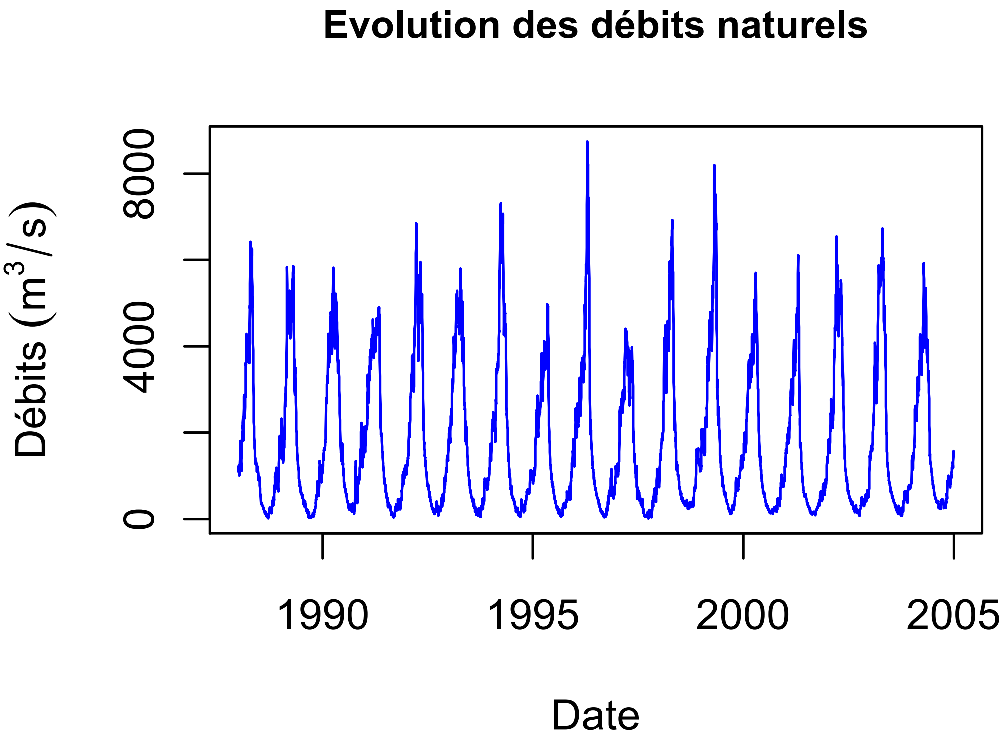

Figure 5.2 – Evolution des débits naturels journaliers reconstitués du bassin versant intermédiaire de la Sanaga à la station de Sogmbengué.

prévisibles), nous allons les extraire de la série. Considérons la série standardisée définie par

$$Y_t = \frac{x_{i,j} - \overline{x}_i}{\widehat{\sigma}_i} \quad \text{lorsque } t \text{ est le jour } i \text{ de l'année } j.$$

Définissons la série differenciée

$$X_t = Y_t - Y_{t-1}.$$

La Figure 5.4 présente une représentation graphique de la serie $(X_t)_{t=1,\cdots,6204}$.

Dkengne 2006 a modélisé la série $(X_t)_t$ par un ARMA(11,1). Vu que les coefficients ar10-ar7 ne sont pas significatifs, nous allons appliquer un modèle ARMA(6,1).

5.2 Detection de rupture par un critère de quasi-vraisemblance pénalisé

Nous allons utiliser la procédure décrite dans Bardet *et al.* 2011 pour étudier l'existence des ruptures dans la série $(X_t)_{t=1,\cdots,n}$ avec $n = 6204$.

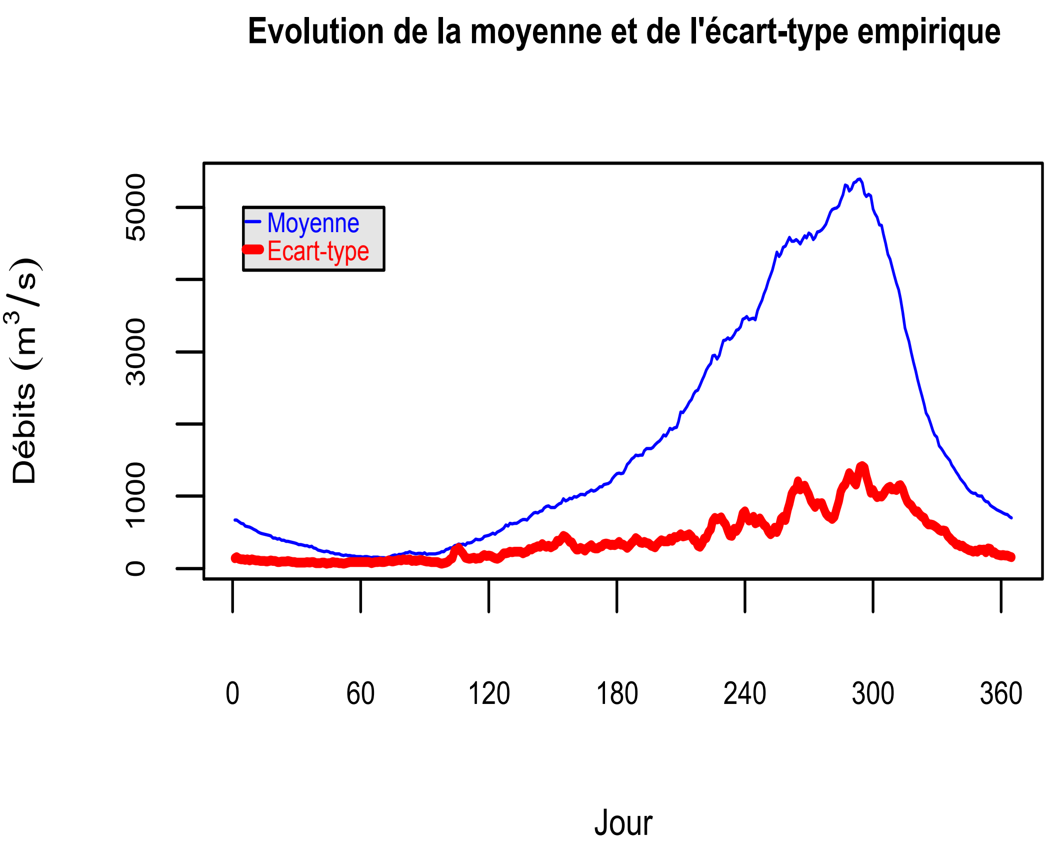

FIGURE 5.3 – Evolution de la moyenne et de l'écart-type empirique des débits naturels journaliers reconstitués du bassin versant intermédiaire de la Sanaga à la station de Sogmbengué.

5.2.1 Le modèle

Comme cela a été mentionné ci-dessus, nous allons appliquer un modèle ARMA(6,1) : $X_t = \sum_{k=1}^{6} \phi_k X_{t-k} + \xi_t + \theta_1 \xi_{t-1}$. Le paramètre du modèle est noté $\theta = (\phi_1, \phi_2, \cdots, \phi_6, \theta_1) \in \Theta$ où Θ est un compact de $\mathbb{R}^7$. Rappelons qu'un ARMA(p,q) appartient à la classe $\mathcal{M}_{\mathbb{Z}}(M, f)$ par la représentation AR(∞). On fixe $K_{max} \leq 14$. Rappelons la définition du critère de vraisemblance non pénalisé

$$(QLIK) \quad \widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2 \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k)$$

et du critère pénalisé

$$(penQLIK) \quad \widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n K$$

où $K \in \{1, 2, \cdots, K_{max}\}$, $\underline{t} = (t_0, t_1, \ldots, t_{K-1}, t_K)$ avec $0 = t_0 < t_1 < \ldots < t_{K-1} < t_K = n$ $\underline{\theta} = (\theta_1, \ldots, \theta_K) \in \Theta^K$, $T_k =]t_{k-1}, t_k] \cap \mathbb{N}$, $\widehat{L}_n(T_k, \cdot)$ la quasi-vraisemblance du modèle calculée sur T_k (voir Bardet *et al.* 2011), $\kappa_n < n$ le paramètre de pénalité.

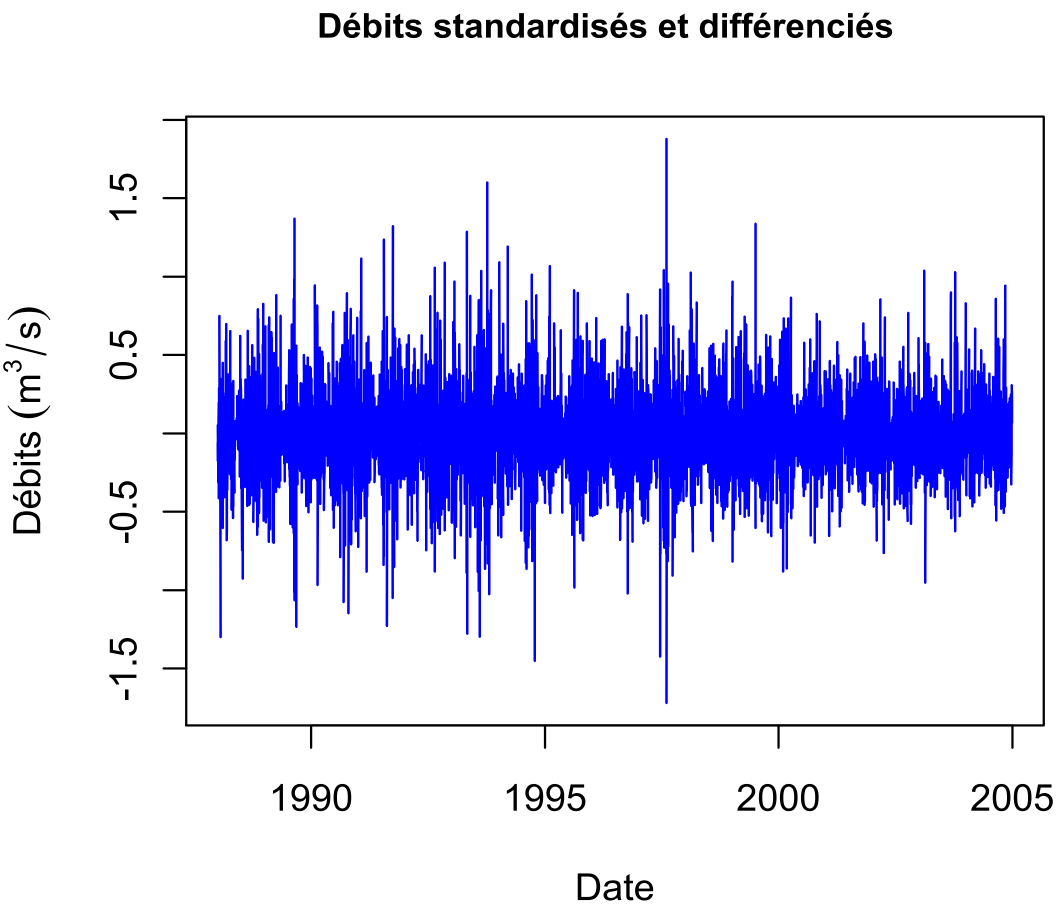

Figure 5.4 – Evolution des débits standardisés et différenciés.

5.2.2 Estimation du paramètre de pénalité par l'heuristique de la pente

Comme dans Bardet *et al.* 2011, nous allons appliquer l'heuristique de la pente pour estimer le paramètre de pénalité κ_n. Rappelons que cette procédure consiste à estimer κ_n par $\widehat{\kappa}_n$ où $\widehat{\kappa}_n/2$ est la pente de la partie linéaire de la courbe $(K, -\min\limits_{\underline{t},\underline{\theta}} QLIK(K))_{1\leq K\leq K_{max}}$. La Figure 5.5 prensente la pente de cette courbe.

Il ressort de la Figure 5.5 que l'estimation du paramètre de pénalité est $\hat{\kappa}_n = 17.5$.

5.2.3 Estimation du modèle

En utilisant $\hat{\kappa}_n = 17.5$, nous allons maintenant minimiser le critère $penQLIK$ en $(K, \underline{t}, \underline{\theta})$, avec $1 \leq K \leq K_{max}$. La Figure 5.6 représente les points $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ pour $1 \leq K \leq K_{max} = 12$.

Ainsi, il ressort de la Figure 5.6 que le nombre de segments estimés est $\widehat{K} = 4$ (i.e. 3 ruptures dans le modèle). Les instants de rupture estimés sont : $\widehat{t}_1 = 2093$, $\widehat{t}_2 = 2821$ et $\widehat{t}_3 = 3409$. La Figure 5.7 montre les instants de ruptures dans la série. Pour apprécier la qualité de ces estimations, notons d'abord que la théorie développée au chapitre 2

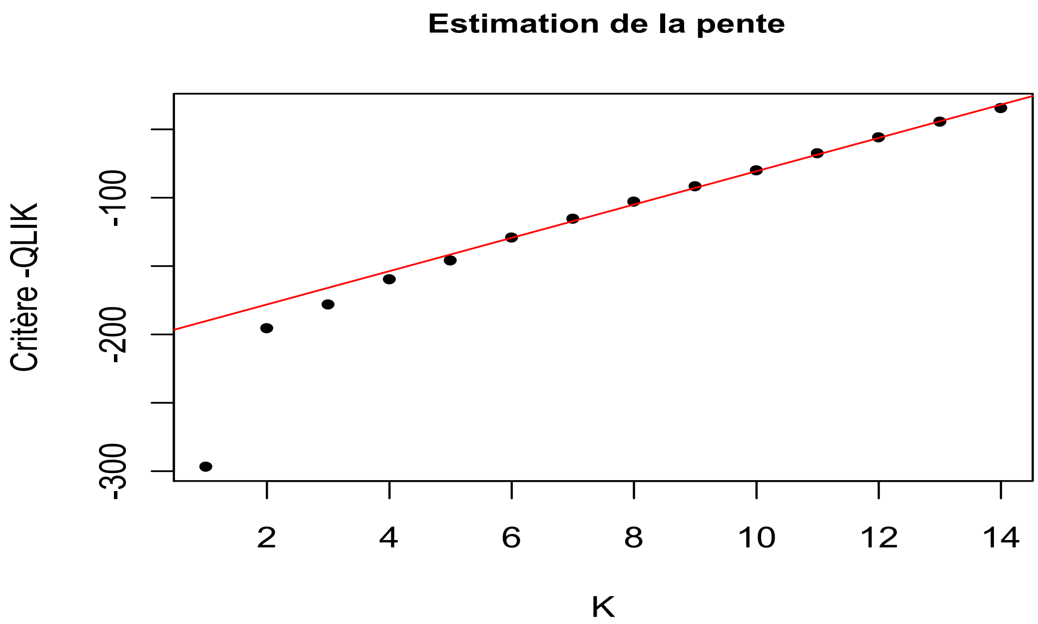

FIGURE 5.5 – La courbe de $-\min_{t,\theta} QLIK$ avec $1 \leq K \leq 14$. La pente de la partie linéaire (de la droite) est $\hat{\kappa}_n/2 = 8.75$.

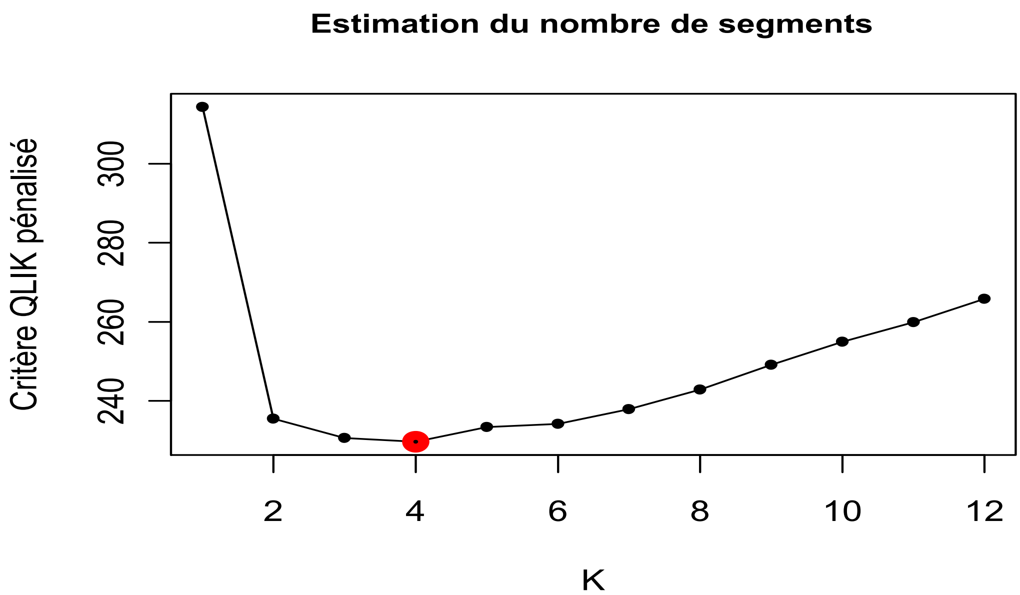

FIGURE 5.6 – Courbe du critère QLIK pénalisé.

montre que le nombre de rupture estimé converge en probabilité vers le vrai nombre de rupture. Les résultats des simulations (toujours dans ce chapitre 2) ont montré que pour une taille d'échantillon $n = 1000$, la probabilité d'estimer le bon nombre de ruptures se situe autour de 0.75. Ces résultats ont aussi montré que cette probabilité augmente de façon considérable avec la taille de l'échantillon. On peut donc conclure qu'il y a une forte probabilité pour que cette série hydrologique comporte 3 ruptures. Ensuite, pour les instants de rupture estimés, on sait (d'après le chapitre 2) que l'écart entre les ruptures estimées et les vraies ruptures est assez faible et augmente très lentement avec n. Ce qui donne de bonnes raisons de penser que ces ruptures estimées dans la série hydrologique sont assez proches des ruptures réelles.

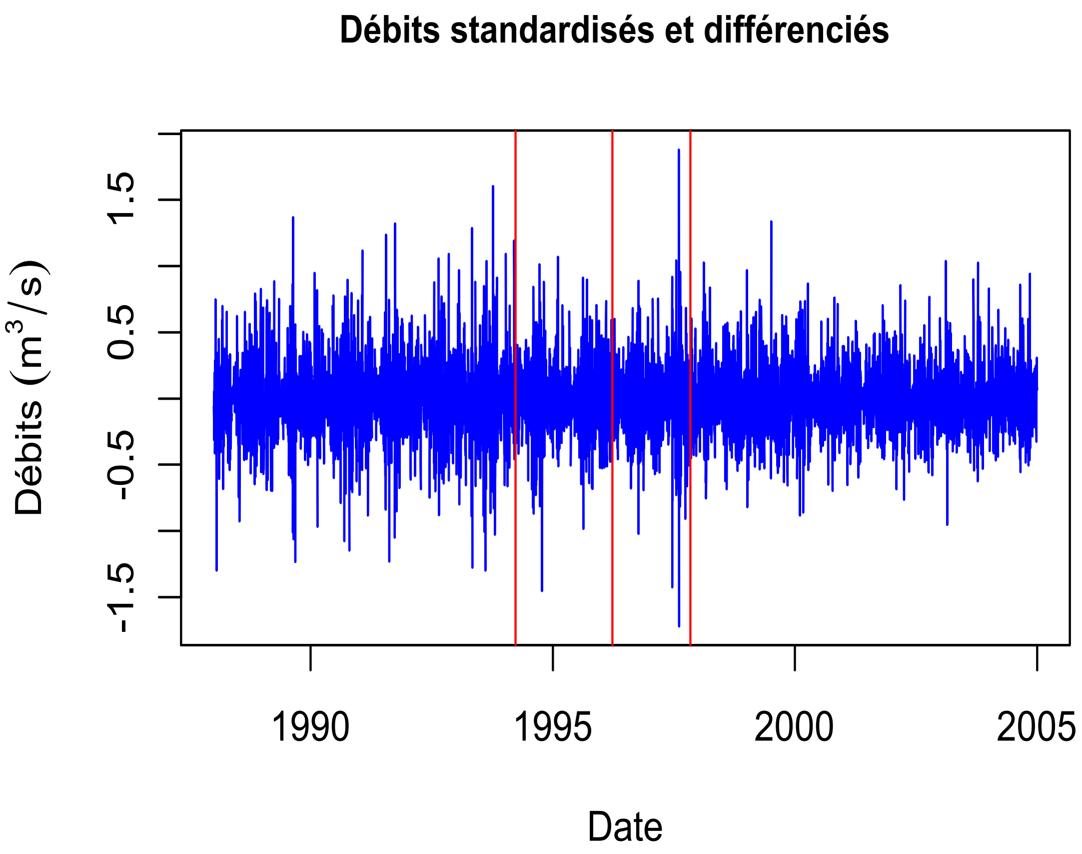

Figure 5.7 – Les droites verticales indiquent les instants de rupture détectés par le critère de vraisemblance pénalisée.

Nous allons maintenant estimer les paramètres du modèle sur chaque segment. Posons $\mathbb{X}_1 = (X_t)_{t=1,\cdots,\widehat{t}_1}$, $\mathbb{X}_2 = (X_t)_{t=\widehat{t}_1+1,\cdots,\widehat{t}_2}$. $\mathbb{X}_3 = (X_t)_{t=\widehat{t}_2+1,\cdots,\widehat{t}_3}$ et $\mathbb{X}_4 = (X_t)_{t=\widehat{t}_3+1,\cdots,6204}$. Pour chacune des séries $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$, on fait une sélection par le critère BIC des ordres p et q de l'ARMA. Le meilleur modèle BIC est cherché avec $(p,q) \in \{1,\cdots,15\} \times \{1,\cdots,15\}$. On otient : $\mathbb{X}_1 \sim ARMA(6,1)$, $\mathbb{X}_2 \sim ARMA(1,1)$, $\mathbb{X}_3 \sim ARMA(1,0)$ et $\mathbb{X}_4 \sim ARMA(5,1)$. Ce qui confirme bien que le modèle ARMA(6,1) utilisé au départ était indiqué pour couvrir tous les sous-modèles de la série. Le Tableau 5.1 donne les paramètres estimés (avec les intervalles de confiance à 95%) du modèle ARMA sur chaque segment (*i.e.* pour chacun des processus $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$). Les distributions des résidus de ces

	$\mathbb{X}_1$	$\mathbb{X}_2$	$\mathbb{X}_3$	$\mathbb{X}_4$
ar_1	0.910	-0.785	0.185	0.970
	$]0.865, 0.955[$	$]-0.919, -0.572[$	$]0.106, 0.265[$	$]0.966, 0.991[$
ar_2	0.109	NA	NA	0.011
	$]0.052, 0.167[$	NA	NA	$]-0.040, 0.064[$
ar_3	-0.122	NA	NA	-0.128
	$]-0.179, -0.064[$	NA	NA	$]-0.181, -0.076[$
ar_4	0.043	NA	NA	0.075
	$]-0.014, 0.101[$	NA	NA	$]0.023, 0.128[$
ar_5	0.066	NA	NA	-0.031
	$]0.008, 0.123[$	NA	NA	$]-0.070, 0.007[$
ar_6	-0.104	NA	NA	NA
	$]-0.148, -0.061[$	NA	NA	NA
ma_1	-0.979	0.722	NA	-0.980
	$]-0.995, -0.963[$	$]0.487, 0.957[$	NA	$]-0.990, -0.965[$

TABLE 5.1 – Estimation des paramètres sur chaque segment.

différents modèles sont données par la Figure 5.8. Le tableau 5.2 donne les résultats du test du Portemanteau avec un décalage $lag = 30$. On voit bien que l'hypothèse selon laquelle ces

	$\mathbb{X}_1$	$\mathbb{X}_2$	$\mathbb{X}_3$	$\mathbb{X}_4$
p-value	0.075	0.10	0.09	0.70

TABLE 5.2 – Test du Portemanteau sur les résidus de $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$.

résidus sont des bruit blancs est validée. Les autocorrélations et autocorrélations partielles de ces résidus sont données par la Figure 5.9. Cette figure montre également que les autocorrélations de ces résidus sont significativement nulles.

Le Figure 5.10 représente les périodogrammes lissés des séries $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$. Les comportements basse, moyenne et haute fréquences de ces spectres montrent bien que les structures des segments détectés sont différentes.

Lavielle [49] a développé une procédure permettant de détecter les ruptures à partir d'un contraste pénalisé (procédure appelée DCPC). Le contraste utilisé est une fonction de la forme

$$J(\underline{t}, X_1, \cdots, X_n) = \frac{1}{n}\sum_{k=1}^{K} G(X_{t_{k-1}+1}, \cdots, X_{t_k})$$

où $\underline{t}$ est une configuration de rupture avec $\underline{t} = (t_0, \cdots, t_K)$, $t_0 = 0$ et $t_K = n$. On pose

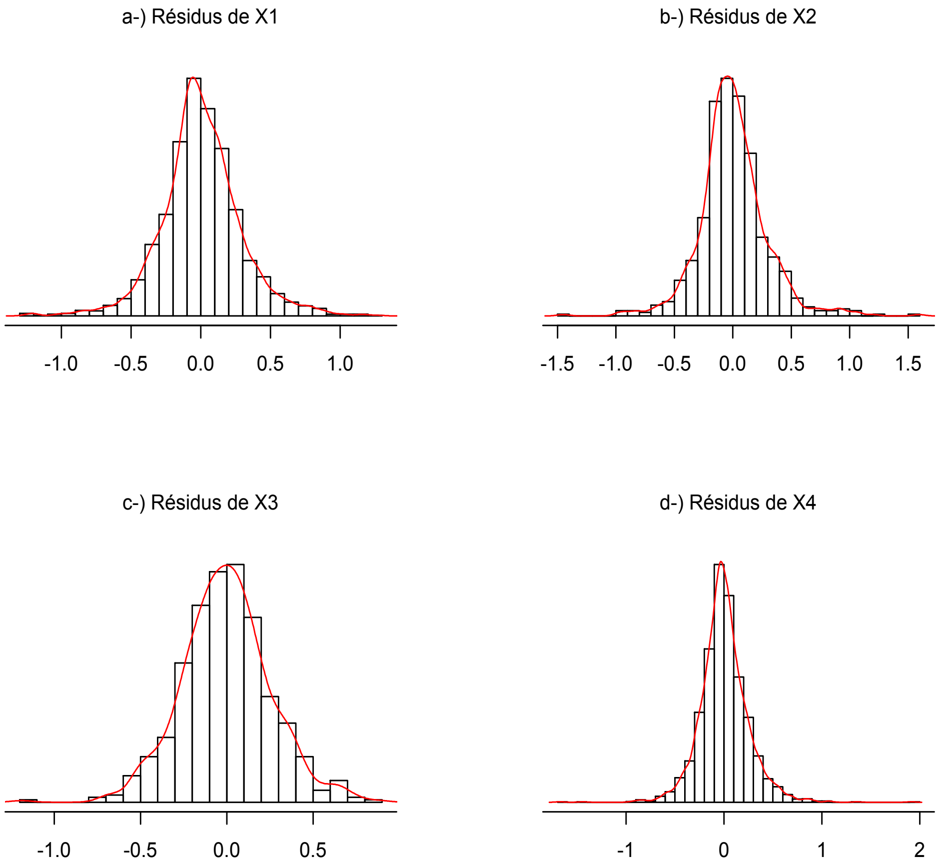

FIGURE 5.8 – Résidus des modèles estimés à partir des séries $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$.

$T_k = \{t_{k-1}+1, t_{k-1}+2, \cdots, t_k\}$ et $n_k = Card(T_k)$. La quantité $G(X_{t_{k-1}+1}, \cdots, X_{t_k})$ dépend du type de rupture que l'on souhaite détecter.

- Pour une détection de rupture dans la moyenne, prendre $G(X_{t_{k-1}+1}, \cdots, X_{t_k}) = \sum_{j \in T_k}(X_j - \overline{X}_{T_k})^2$ où $\overline{X}_{T_k} = \frac{1}{n_k}\sum_{j \in T_k} X_j$.
- Pour une détection de rupture dans la variance, prendre $G(X_{t_{k-1}+1}, \cdots, X_{t_k}) = n_k \log(\widehat{\sigma}^2_{T_k})$ avec $\widehat{\sigma}^2_{T_k} = \frac{1}{n_k}\sum_{j \in T_k}(X_j - \overline{X}_n)^2$.
- Pour une détection de rupture dans la densité spectrale, poser $I_k(\lambda) = \frac{1}{2\pi n_k}|\sum_{j \in T_k} X_j e^{-ij\lambda}|^2$ avec $\lambda \in [0, \pi]$; $F_k = \int_{-\pi}^{\pi} I_k(\lambda) d\lambda$ et prendre $G(X_{t_{k-1}+1}, \cdots, X_{t_k}) = -n_k F_k^2$.

Le terme de pénalité de la méthode DCPC est de la forme $\beta pen(K)$. Lavielle développe une procédure automatique permettant d'estimer le paramètre β de manière adaptative. Cette procédure est basée sur la recherche de la plus grande courbure de la courbe $(pen(K), J_K)$ où J_K est le contraste calculé sur le modèle à K segments. Voir [49] pour plus de connaissance sur cette procédure. Les Figures 5.11 5.12 donnent les résultats d'application de la procédure DCPC sur les débits hydrologiques hebdomadaires.

On peut voir sur la Figure 5.11 a-) qu'aucune rupture n'a été détectée dans la moyenne. Les ruptures détectées dans la variance correspondent aux instants $t_1 = 2604$ et $t_2 = 2800$ dans la série. Les ruptures détectées dans le périodogramme correspondent aux instants $t_3 = 1946$ et $t_4 = 2114$. Les ruptures t_2 et t_4 ont été détectées par la procédure QLIK.

Ce qui montre bien que la procédure QLIK est susceptible de capter le changement dans la variance et dans le périodogramme simultanément. Notons que la procédure DCPC est fondée sur le fait qu'il y a stationnarité sur les segments même après la rupture ; cette hypothèse est généralement violée sur les données hydrologiques.

5.3 Detection de rupture par une procédure de test

Nous allons appliquer la procédure de test proposée par Kengne 2011 et l'algorithme ICSS proposé par Inclan et Tiao 1994 (voir aussi Sous-section 1.4.4) pour détecter les ruptures dans les observations $(X_t)_{t=1,\cdots,6204}$. Nous allons comme précedemment appliquer un modèle ARMA(6,1). Nous calculerons la statistique $\widehat{Q}_n := \max(\widehat{Q}_n^{(1)}, \widehat{Q}_n^{(2)}) = \max_{k\in\Pi_n}\Big(\max(\widehat{Q}_{n,k}^{(1)}, \widehat{Q}_{n,k}^{(2)})\Big)$ (voir Sous-section 1.4.2) et comparerons à C_α, le quantile d'ordre $(1-\alpha/2)$ de la loi de $\sup_{0\leq\tau\leq 1}\|W_7(\tau)\|^2$ où W_7 est un pont Brownien de dimension 7. Pour $\alpha = 0.05$, un calcul par la méthode de Monte Carlo donne $C_\alpha = 5.26$.

1. **Etape 0** : on pose $t_0 = 0$.
2. **Etape 1** : On fait le test avec toutes les observations $(X_t)_{t=1,\cdots,6204}$. La Figure 5.13 donne une représentation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$.
   On déduit de la Figure 5.13 un point de rupture potentiel en $k = 3504$. Posons $t_1 = 3504$.
3. **Etape 2**
   - **Etape 2a** : on cherche le premier instant de rupture.
     i-) On fait le test avec les observations $(X_t)_{t=1,\cdots,3504}$. La Figure 5.14 donne une représentation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$. On déduit de la Figure 5.14 un point de rupture potentiel en $k = 2117$. Posons donc $t_1 = 2117$.
     ii-) On fait le test avec les observations $(X_t)_{t=1,\cdots,2117}$. La Figure 5.15 donne une représentation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$.
     On déduit de la Figure 5.15 un point de rupture potentiel en $k = 1621$. Posons de nouveau $t_1 = 1621$.
     iii-) On fait le test avec les observations $(X_t)_{t=1,\cdots,1621}$. La Figure 5.16 donne une représentation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$.
     D'après la Figure 5.16, il n'y a pas de rupture dans les observations $(X_t)_{t=1,\cdots,1621}$.
     On déduit donc que le premier instant de rupture potentiel est $t_{first} = t_1 = 1621$.
   - **Etape 2b** : On procède comme à l'**Etape 2a** pour chercher le dernier instant de rupture en partant des observations $(X_t)_{t=1622,\cdots,6204}$. On trouve $t_{last} = 4596$.
   - **Etape 2c** : On reprend les Etape 2a et Etape 2b sur les observations $X_{1622}, \cdots, X_{4596}$. On trouve trois nouveaux points de rupture potentiels qui sont $t_2 = 2117$, $t_3 = 2833$ et $t_4 = 3453$.
4. **Etape 3** : On teste que les points $t_{first} = t_1 = 1621$, $t_2 = 2117$, $t_3 = 2833$, $t_4 = 3453$ et $t_{last} = 4596$ sont effectivement des instants de rupture. On trouve finalement les instants de ruptures estimés : $\widehat{t}_1 = 1621$, $\widehat{t}_2 = 2117$, $\widehat{t}_3 = 2833$ et $\widehat{t}_4 = 3453$.

La Figure 5.17 montre ces instants de ruptures dans la série. On remarque que l'instant de rupture $\widehat{t}_1 = 1621$ n'a pas été détecté par la méthode de vraisemblance pénalisée. Partant de cette rupture, on obtient les séries $\mathbb{X}_1(1) = (X_t)_{t=1,\cdots,1621}$ et $\mathbb{X}_1(2) = (X_t)_{t=1622,\cdots,2117}$. En faisant l'identification de ces deux séries par un critère BIC, on trouve que $\mathbb{X}_1(1) \sim ARMA(6,1)$ et $\mathbb{X}_2(2) \sim ARMA(0,1)$. Le Tableau 5.3 donne les estimations des paramètres de ces deux modèles.

	$\mathbb{X}_1(1)$	$\mathbb{X}_1(2)$
ar_1	0.957	NA
	$]0.916, 0.981[$	NA
ar_2	0.087	NA
	$]0.019, 0.154[$	NA
ar_3	-0.162	NA
	$]-0.230, -0.095[$	NA
ar_4	0.041	NA
	$]-0.025, 0.109[$	NA
ar_5	0.080	NA
	$]0.012, 0.147[$	NA
ar_6	-0.111	NA
	$]-0.161, -0.062[$	NA
ma_1	-0.967	-0.260
	$]-0.985, -0.961[$	]-0.347,-0.173[

TABLE 5.3 – Estimation des paramètres des séries $\mathbb{X}_1(1)$ et $\mathbb{X}_1(2)$.

La différence entre ces deux modèles est significative et confirme ainsi le fait qu'il y a un point de rupture en $\widehat{t}_1 = 1621$.

Ainsi, les débits standardisés et différenciés, du bassin versant de la Sanaga sont découpés en cinq segments représentés par les séries $\mathbb{X}_1(1)$, $\mathbb{X}_1(2)$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$. Les estimations des paramètres montrent des changements dans la structure de la série. Ces changements affectent les variances des régimes stationnaires et induisent de très fortes variations dans le débit réel. Après des discussions menées avec des Hydrologues et des Climatologues, notre avis est que ces ruptures détectées sont dues aux changements climatiques.

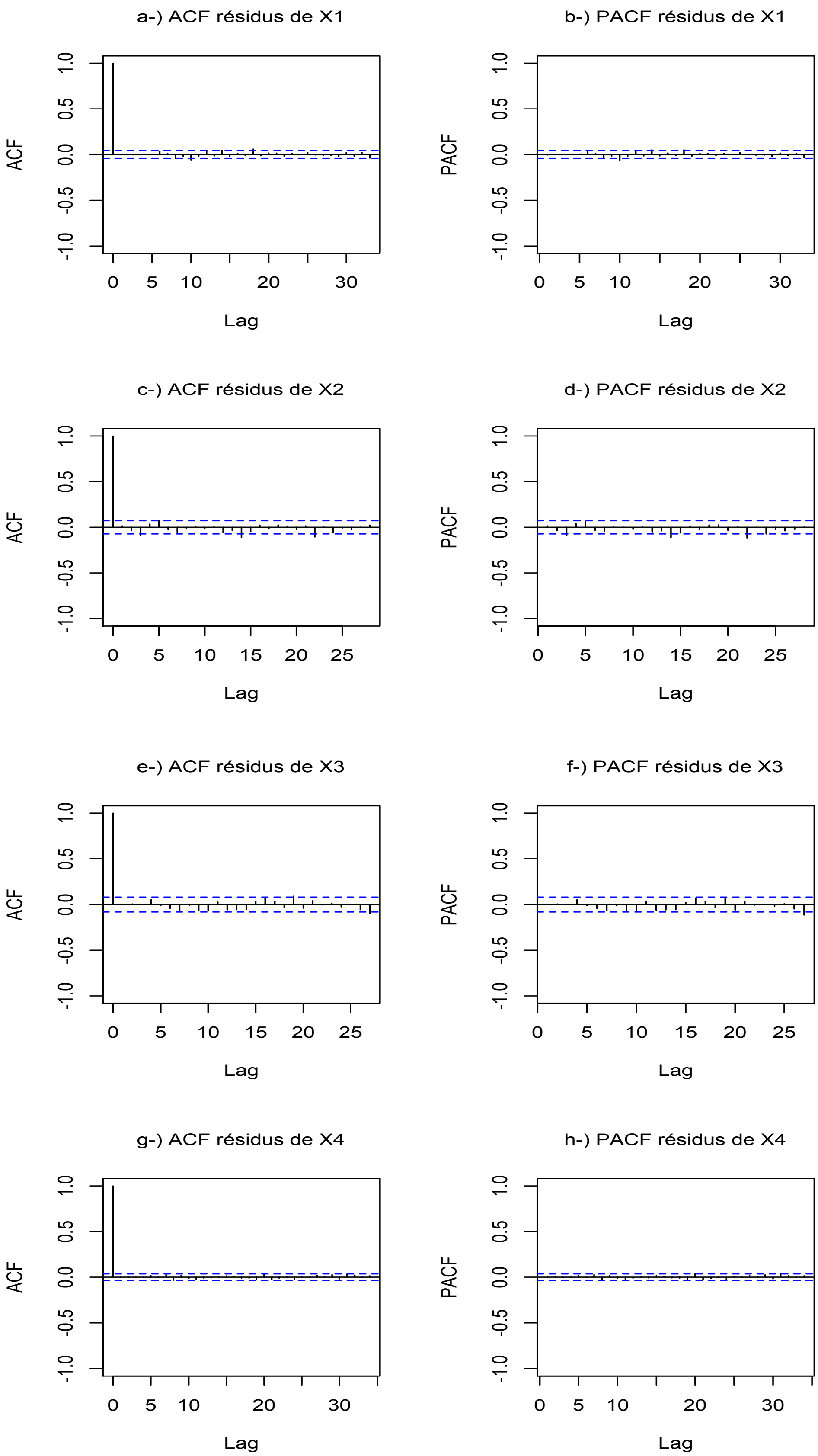

Figure 5.9 – Autocorrélations et autocorrélations partielles des résidus issus de $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$.

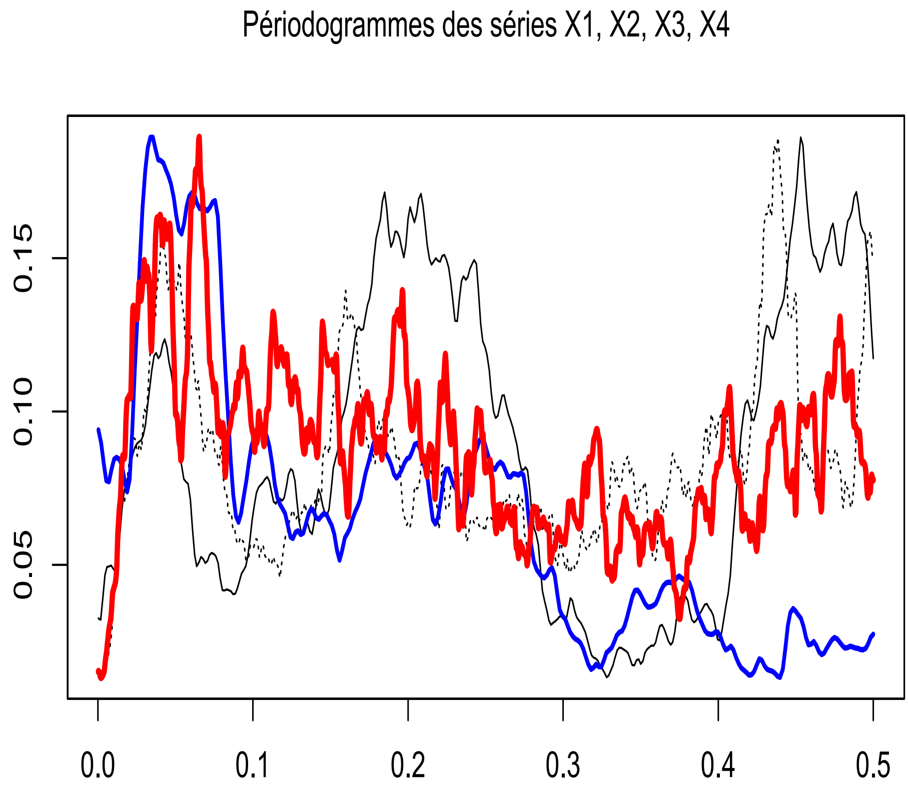

FIGURE 5.10 – Périodogrammes de $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$. Les traits interrompus, continus, moyens et très forts représentent les périodogrammes des séries $\mathbb{X}_1$, $\mathbb{X}_2$, $\mathbb{X}_3$ et $\mathbb{X}_4$ respectivement.

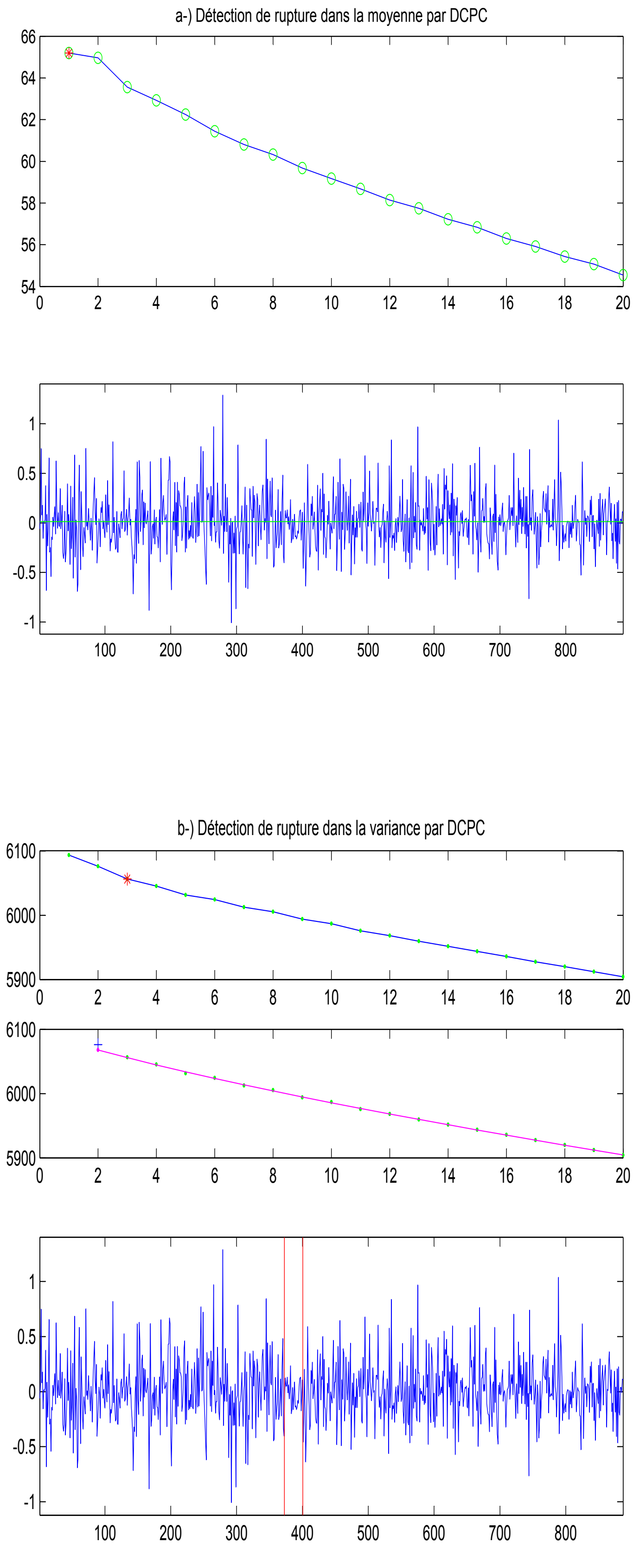

FIGURE 5.11 – Détection de rupture dans la moyenne et dans la variance par la procedure DCPC.

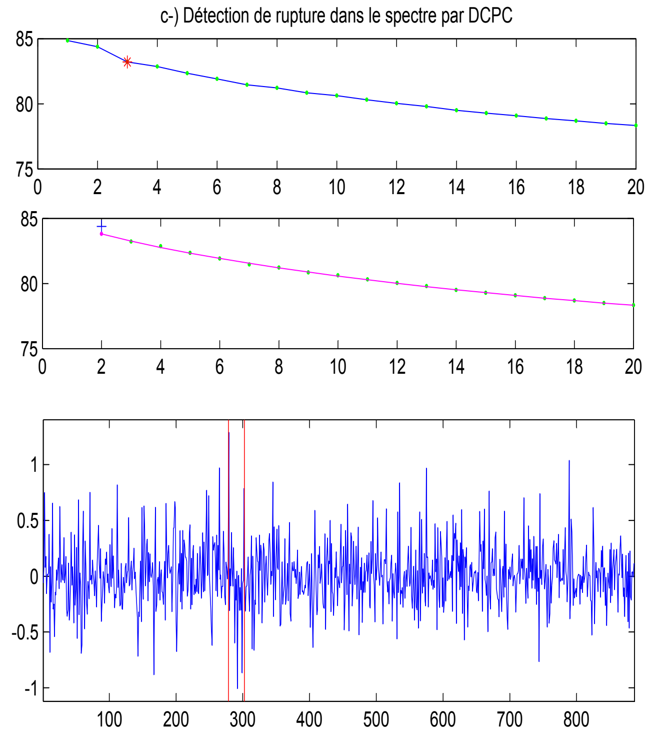

FIGURE 5.12 – Détection de rupture dans le spectre par la procedure DCPC.

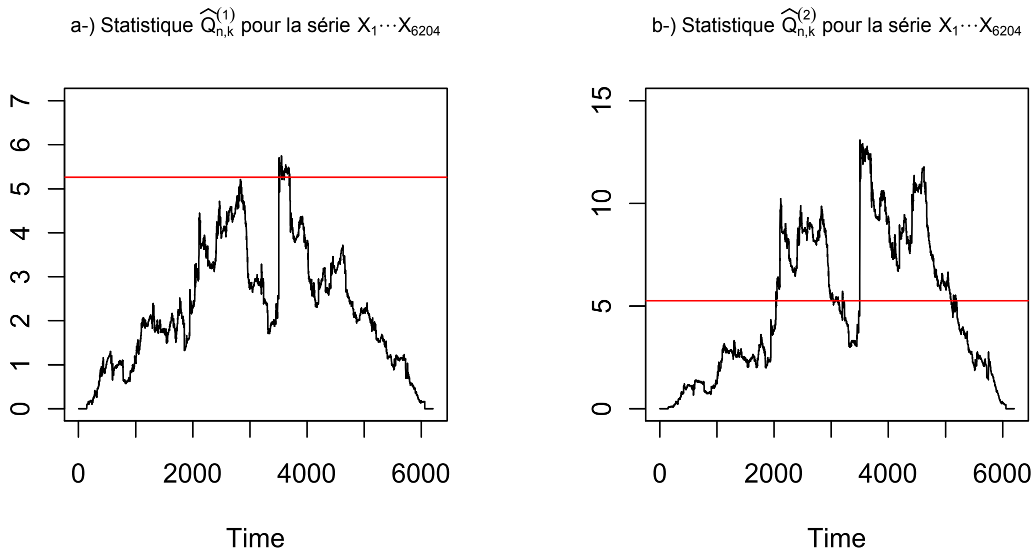

FIGURE 5.13 – Réalisation des statistiques $\widehat{Q}^{(1)}_{n,k}$ et $\widehat{Q}^{(2)}_{n,k}$ pour les observations $(X_t)_{t=1,\cdots,6204}$. La droite horizontale représente la valeur critique du test.

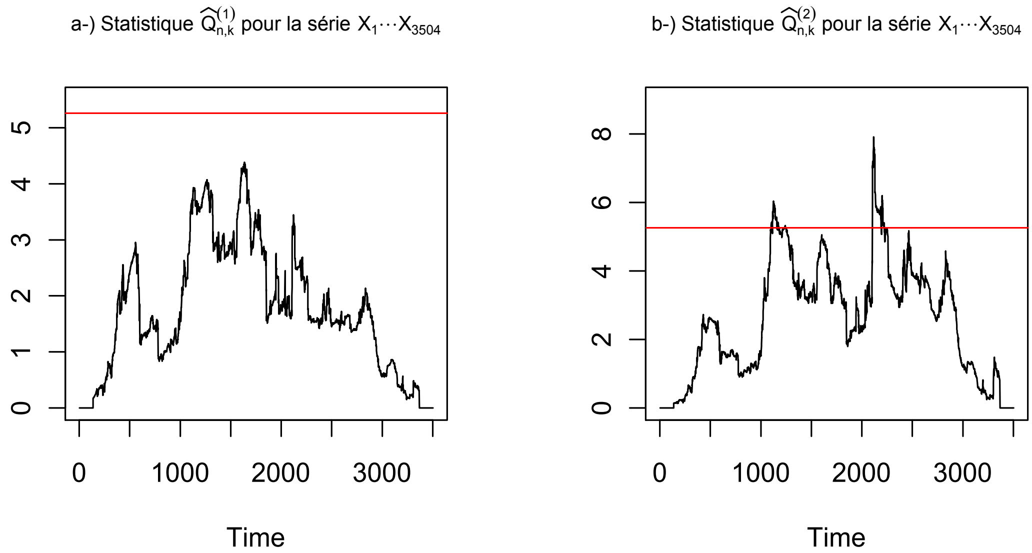

FIGURE 5.14 – Réalisation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$ pour les observations $(X_t)_{t=1,\cdots,3504}$. La droite horizontale représente la valeur critique du test.

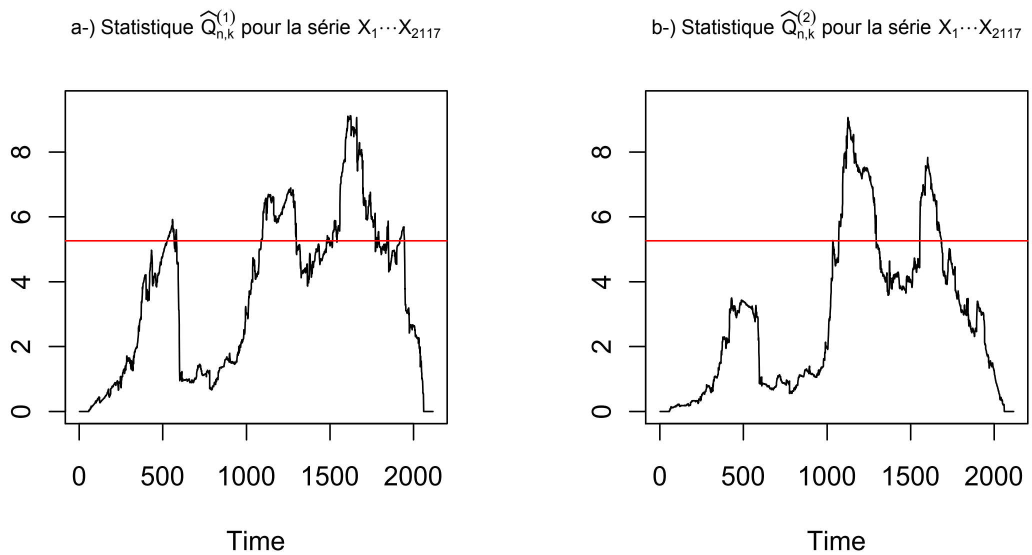

FIGURE 5.15 – Réalisation des statistiques $\widehat{Q}_{n,k}^{(1)}$ et $\widehat{Q}_{n,k}^{(2)}$ pour les observations $(X_t)_{t=1,\cdots,2117}$.

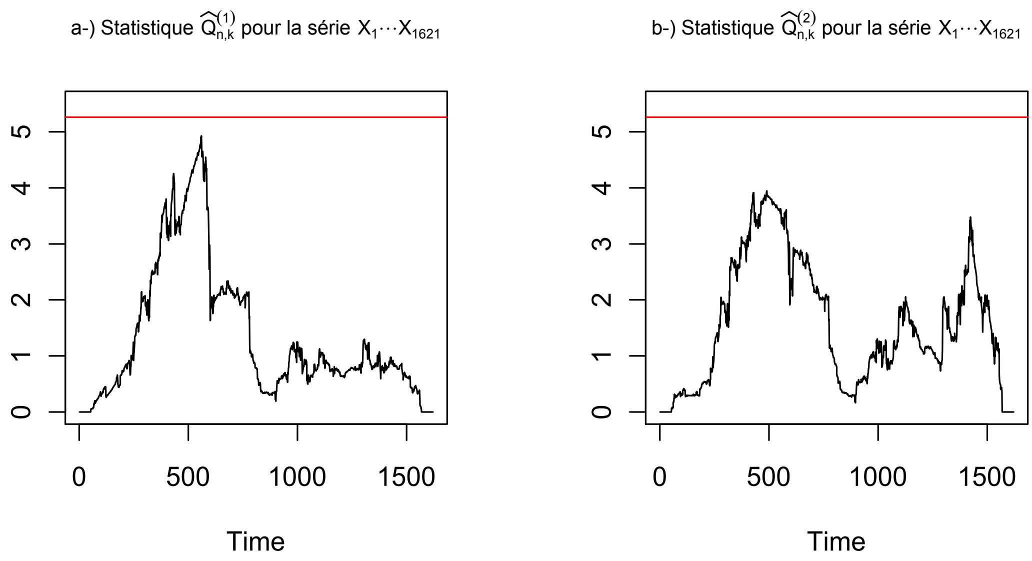

Figure 5.16 – Réalisation des statistiques $\widehat{Q}^{(1)}_{n,k}$ et $\widehat{Q}^{(2)}_{n,k}$ pour les observations $(X_t)_{t=1,\cdots,1621}$.

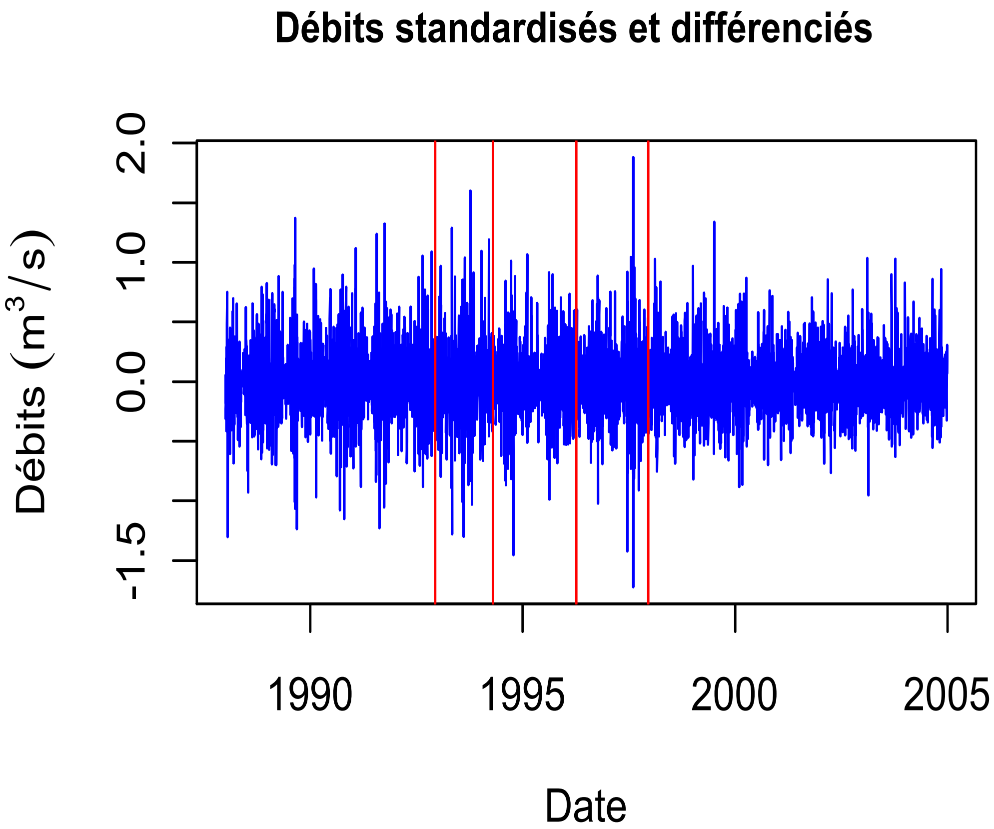

FIGURE 5.17 – Les droites verticales indiquent les instants de rupture détectés par par le procédure de test.

Chapitre 6

Conclusion et perspectives

Cette thèse porte sur la détection de rupture dans les processus causaux. Les détections off-line et on-line sont traitées dans un cadre semi-paramétrique. Les procédures développées sont évaluées à travers des simulations et les résultats sont satisfaisants. Elles sont donc appliquées aux débits du bassin versant de la Sanaga. Les résultats montrent la présence des ruptures structurelles dans ces débits.

Dans un premier temps, nous présentons le problème de rupture, la classe semi-paramétrique à étudier et donnons des exemples de modèles classiques usuels.

Ensuite, nous développons une procédure de détection off-line de rupture multiple, basée sur un critère de vraisemblance pénalisée. Bien que le processus a perdu sa stationnarité après le premier instant de rupture, nous démontrons qu'on peut toujours l'approcher (sur chaque segment) de manière efficace par son régime stationnaire. Ce résultat permet de montrer que les estimateurs du modèle de rupture convergent avec des vitesses optimales. Un estimateur adaptatif du paramètre de pénalité basé sur l'heuristique de la pente est proposé. Cette procédure est assez robuste et assez efficace comme le montrent les résultats des simulations et les résultats de l'application à l'indice FTSE

Toujours dans le cadre off-line, nous développons une nouvelle procédure permettant de tester la présence de rupture dans les processus causaux. La consistance en puissance est prouvée. Les résultats de simulations confirment l'applicabilité de la procédure. Avec un algorithme de type ICSS, cette procédure permet de détecter les ruptures multiples. Nous conseillons l'utilisation de cette procédure lorsque la taille de l'échantillon est assez-grande ($n > 2500$), car dans ce cas, la procédure basée sur le critère pénalisé devient numériquement difficile à mettre en oeuvre.

La détection on-line de rupture est traitée au chapitre 4. Nous proposons un nouveau détecteur construit à partir des quasi-vraisemblances des observations. Le comportement asymptotique de ce détecteur est étudié. Le résultat obtenu permet de montrer que la procédure on-line basée sur ce détecteur est consistante. Des résultats de simulations montrent que cette procédure réduit significativement le délai de détection par rapport aux procédures existantes.

Toutefois, dans les procédures développées, la loi asymptotique des ruptures est inconnue. Il est important d'étudier ce comportement asymptotique car il permettra de déterminer les intervalles de confiance et de mieux apprécier la qualité des estimations.

D'autre part, cette thèse s'est fait dans un cadre semi-paramétrique *i.e.* on a supposé que les formes des fonctions f_θ et M_θ sont connues. Une question intéressante serait de savoir ce qui se passe dans un cadre non-paramétrique *i.e.* les fonctions f_θ et M_θ sont inconnues. L'idée naturelle serait d'estimer ces fonctions et d'évaluer l'écart entre la fonction avant la rupture et la fonction après la rupture. Problème : les fonctions f_θ et M_θ ont une infinité de variables. Une très bonne piste pour poursuivre ce travail...

Annexe

Lexique de quelques termes techniques

- **Bassin versant d'un fleuve** : est la totalité de la surface topographique drainée par un fleuve et ses affluents ;

- **Débit d'un cours d'eau** : est le volume total d'eau qui s'écoule à travers une section droite du cours d'eau pendant l'unité de temps. Il est généralement exprimé en m^3/s ;

- **Débit naturel** : le débit d'un cours d'eau est dit naturel lorsque celui-ci n'est pas perturbé par la présence de retenues d'eau artificielles ;

- **Débit naturel reconstitué** : lorsque le fleuve est équipé des retenues artificielles, le débit naturel n'est plus mesurable. Il est estimé et on parle donc de débit naturel reconstitué.

Table des figures

Liste des tableaux

Bibliographie

[1] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. Break detection in the covariance structure of multivariate time series models. *Ann. Statist. 37(6B)*, (2009), 4046-4087.

[2] Aue, A., Horváth, L., Husková, M., and Kokoszka, P. Changepoint monitoring in linear models. *Econometrics Journal 9*, (2006), 373-403.

[3] Aue, A., Horváth, L., Hušková, M. and Ling, S. On distinguishing between random walks and changes in the mean alternatives. *Econometric Theory 25*, (2009), 411-441.

[4] Aue, A., Horváth, L. and Reimherr, M. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics 149*, (2009), 174-190.

[5] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. Break detection in the covariance structure of multivariate time series models. *Ann. Statist. 37(6B)*, (2009), 4046-4087.

[6] Bai, J. On the partial sums of residuals in autoregressive and moving average models. *Journal of Time Series Analysis 14*, (1993) 247–260.

[7] Bai J. and Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica 66* (1998), 47–78.

[8] Bardet, J.-M. and Wintenberger, O. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist. 37*, (2009), 2730–2759.

[9] Bardet, J.-M. , Kengne, W. and Wintenberger, O. Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics 6*, (2012), 435-477.

[10] Bardet, J.-M. and Kengne, W. Monitoring procedure for parameter change in causal time series. To appear (2012+).

[11] Basseville, M. and Nikiforov, I. Detection of Abrupt Changes : Theory and Applications. Prentice Hall, Englewood Cliffs, NJ, 1993.

[12] Baudry, J.-P., Maugis, C. and Michel, B. Slope Heuristics : overview and implementation. *RR-INRIA*, (2010) , n°7223.

[13] Berkes, I., Horváth, L., and Kokoszka, P. GARCH processes : structure and estimation. *Bernoulli 9* (2003), 201–227.

[14] Berkes, I., Gombay E., Horváth, L., and Kokoszka, P. Sequential change-point detection in GARCH(p,q) models. *Econometr Theory 20*, (2004), 1140-1167.

[15] Berkes, I., Horváth, L. and Kokoszka, P. Testing for parameter constancy in GARCH(p ; q) models. *Statistics* & *Probability Letters 70*, (2004), 263–273.

[16] BERKES, I., HORVÁTH, L., KOKOSZKA, P. AND SHAO, Q.-M. Almost sure convergence of the Bartlett estimator.. *Period Math Hung 51* (2005), 11–25.

[17] BERKES, I., HORVÁTH, L., KOKOSZKA, P.S. AND SHAO, Q.-M. On discriminating between long-range dependence and changes in mean. *Ann. Statist. 34*, (2006), 1140-1165.

[18] BILLINGSLEY, P. *Convergence of Probability Measures.* John Wiley & Sons Inc., New York, 1968.

[19] BIRGE, L. AND MASSART, P. Minimal penalties for Gaussian model selection. *Probab. Theory Related Fields, 138(1-2)*, (2007) 33–73.

[20] BOLLERSLEV, T. *Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31*, (1986), 307-327.

[21] BROWN, R.L., DURBIN, J., EVANS, J.M.,. *Techniques for testing the constancy of regression relationships over time. Journal of Royal Statistical Society B, 37*, (1975), 149-192.

[22] CHOW, Y.S. *A martingale inequality and the law of large numbers. Proceedings of the American Mathematics Society , 11*, (1960), 107-111.

[23] CHU, C-SJ., STINCHCOMBE, M., AND WHITE, W. *Monitoring structural change*, Econometrica, 64, (1996), 1045-1065.

[24] DAVIS, R. A., HUANG, D. AND YAO, Y.-C. Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist. 23*, (1995), 282–304.

[25] DAVIS, R. A., LEE, T. C. M. AND RODRIGUEZ-YAM, G. A. Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis 29*, (2008), 834–867.

[26] DKENGNE, S. P. Modelisation et prevision des debits naturels journaliers du b.v.i. de la sanaga a la station de contrôle de songmbengue. *Mémoire de Master de statistique, ENSP, Yaoundé*, (2006).

[27] DOUKHAN, P. AND WINTENBERGER, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl. 118*, (2008) 1997-2013.

[28] DUFLO, M. Méthodes récursives aléatoires. *Masson, Paris*, 1990 ; English Edition, Springer, 1996.

[29] FRANCQ, C., AND ZAKOÏAN, J.-M. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10* (2004), 605–637.

[30] FRYZLEWICZ, P., AND SUBBA RAO, S. BaSTA : consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes. *Preprint* (2010) `url:http://stats.lse.ac.uk/fryzlewicz/basta/basta.pdf`

[31] HINKLEY, D. V. Inference about the change-point in a sequence of random variables. *Biometrika 57*, (1970), 1–17.

[32] HORVÁTH, L. Change in autoregressive processes. *Stochastic Processes. Appl. 44*, (1993), 221-242.

[33] HORVÁTH, L., HORVÁTH, Z. AND HUSKOVÁ, M. Ratio tests for change point detection. *Inst. Math. Stat. 1*, (2008), 293-304.

[34] HORVÁTH, L., HUŠKOVÁ, M., KOKOSZKA, P., STEINEBACH, J. Monitoring changes in linear models. *Journal of Statistical Planning and Inference 126*, (2004), 225-251.

[35] HUŠKOVÁ, M., PRÁŠKOVÁ, Z., STEINEBACH, J. On the detection of changes in autoregressive time series I. Asymptotics. *J. Statist. Plann. Inference 137*, (2007), 1243-1259.

[36] Hušková, M., Kirch, C., Prášková, Z., Steinebach, J. On the detection of changes in autoregressive time series, II. Resampling procedures. *J. Statist. Plann. Inference 138*, (2008), 1697-1721.

[37] Inclán, C. and Tiao, G. C. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association 89*, (1994), 913–923.

[38] Jeantheau, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14* (1998), 70–86.

[39] Kay, S. M. Fundamentals of Statistical Signal Processing, vol. 2. *Prentice-Hall, Englewood Cliffs, NJ*, (1998)

[40] Kengne W. Testing for parameter constancy in general causal time-series models. *J. Time Ser. Anal. 33*, (2012), 503-518.

[41] Kiefer, J. K-sample analogues of the Kolmogorov-Smirnov and Cramér-v.Mises tests . *Ann. Math. Statist 30*, (1959), 420–447.

[42] Kim, S., Cho, S. and Lee, S. On the CUSUM test for parameter changes in GARCH(1, 1) models. *Comm. Statist. Theory Methods 29*, (2000), 445-462.

[43] Klimko, L.A. and Nelson, P.I. On conditional least squares estimation for stochastic processes. *Ann. Statist. 6*, (1978), 629-642.

[44] Kokoszka, P. and R. Leipus Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal 39*, (1999), 182-195.

[45] Kokoszka, P. and Leipus, R. Change-point estimation in ARCH models. *Bernoulli 6*, (2000), 513–539.

[46] Kouamo, O., Moulines, E. and Roueff, F. Testing for homogeneity of variance in the wavelet domain. *In Dependence in Probability and Statistics*, P. Doukhan, G. Lang, D. Surgailis and G. Teyssiere. Lecture Notes in Statistic 200 (2010), Springer-Verlag, pp. 420–447.

[47] Kounias, E. G. and Weng, T.-S. An inequality and almost sure convergence. *Annals of Mathematical Statistics 40*, (1969), 1091–1093.

[48] Kulperger R. and Yu, H. High moment partial sum processes of residuals in GARCH models and their applications. *Ann. Statist. 33*, (2005), 2395-2422.

[49] Lavielle, M. Using penalized contrasts for the change-point problem. *Signal Processing 85(8)*, (2005) 1501–1510.

[50] Lavielle, M. and Ludena, C. The multiple change-points problem for the spectral distribution. *Bernoulli 6*, (2000) 845–869.

[51] Lavielle, M. and Moulines, E. Least squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis 21*, (2000) 33–59.

[52] Lebarbier, E. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing 85(4)*, (2005) 717–736.

[53] Lee, S. , HA, J. , and NA, O. The Cusum Test for Parameter Change in Time Series Models . *Scand. J. Statist. 30*, (2003), 781–796.

[54] Lee, S. and NA, O. Test for parameter change in stochastic processes based on conditional least-squares estimator. *J. Multivariate Anal. 93*, (2005), 375-393.

[55] Lee, S., Song, J. Test for parameter change in ARMA models with GARCH innovations . *Statistics & Probability Letters 78*, (2008), 1990–1998.

[56] LEISCH, F., HORNIK, K., AND KUAN, C-H. Monitoring structural changes with the generalized fluctuation test. *Econometr Theory 16*, (2000), 835-854.

[57] MOUSTAKIDES, G. V. sequential change detection revisited sequential change detection revisited . *Ann. Statist. 36(2)*, (2008), 787-807.

[58] NA, O. , LEE, Y. AND LEE, S. Monitoring parameter change in time series models. *Stat Methods Appl 20*, (2011), 171–199.

[59] NDAM, J.R. , KPOUMIE, A. , BOEGLIN, J.L. , LIENOU, G. , NFOCGO, A.K. AND EKODECK, G.E. Transports solides et érosion mécanique dans un écosystème tropical d'Afrique : exemple du bassin versant de la Sanaga au Sud -Cameroun. *Actes des JSIRAUF, Hanoi*, (2007).

[60] NELSON, D. B. AND CAO, C. Q. Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics 10*, (1992), 229–235.

[61] PAGE, E. S. A test for a change in a parameter occurring at an unknown point. *Biometrika 42*, (1955), 523–526.

[62] RABEMANANJARA, R. AND ZAKOÏAN, J. M. Threshold ARCH models and asymmetries in volatility. *Journal of Applied Econometrics 8*, (1993) 31–49.

[63] ROBBINS, M., GALLAGHER, C., LUND, R., AND AUE, A. Mean shift testing in correlated data. *Journal of Time Series Analysis 32*, (2011) 498–511.

[64] ROBINSON, P. M. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *J. Econometrics 47*, (1991), 67–84.

[65] ROBINSON, P., AND ZAFFARONI, P. Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist. 34* (2006), 1049–1074.

[66] SHAO, X., AND ZHANG, X. Testing for Change Points in Time Series. *Journal of the American Statistical Association 105*, (2010), 1228–1240.

[67] SIGHOMNOU, D. Analyse et redefinition des regimes climatiques et hydrologiques du cameroun : Perspectives d'evolution des ressources en eau . *Thèse de Doctorat d'Etat ès-Sciences Naturelles, Université de Yaoundé I*, (2004).

[68] STRAUMANN, D., AND MIKOSCH, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach. *Ann. Statist. 34* (2006), 2449–2495.

[69] YAO, Y.C. Estimating the number of change-points via Schwarz criterion. *Statistics & Probability Letters 6*, (1988), 181–189.

[70] ZEILEIS, A., LEISCH F., KLEIBER, C., AND HORNIK, K. Monitoring structural change in dynamic econometric models. *J. Appl. Econometrics 20*, (2005), 99-121.